

**BEST AVAILABLE COPY**

<http://www19.ipdl.ncipi.go.jp/PA1/result/detail/main/wAAAcZaOvIDA414123837...> 2006/04/28

rejection]

[Date of requesting appeal against examiner's decision of  
rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

## \* NOTICES \*

JPO and NCIP1 are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

## CLAIMS

## [Claim(s)]

[Claim 1] The code book which has stored the data which define the 1st face expression set of the 1st person is defined. The data which define the 2nd face expression set which gives a training expression set of the 2nd different person from the 1st person are prepared. Said training expression set, The face expression conversion approach characterized by drawing a transform function from the expression which corresponds from said 1st expression set, and obtaining a synthetic expression set with the application of said transform function to this 1st expression set.

[Claim 2] It is the approach characterized by said training expression set containing an expression fewer than a code book in an approach according to claim 1.

[Claim 3] It is the approach characterized by said transform function compensating the size of the face of the 1st person and the 2nd person, and the difference in a configuration in an approach according to claim 1.

[Claim 4] The approach characterized by calculating the linearity-conversion to the expression set of another side from one expression set in an approach according to claim 1 at the step which draws said transform function.

[Claim 5] In an approach according to claim 1, at the step which draws said transform function Each expression is expressed as a 3m vector containing x in the correlation sample location of m pieces, y, and z variation rates. n expression vector sets of the face changed are ga1... The vector set to which it is given to as n and a target face corresponds is gb1... When given as n, they are 3m linearity least square system of the following format, and [Equation 1].

$$a_j \cdot g_a = g_b[j], i = 1 \dots n$$

The approach characterized by calculating one linear prediction child set aj for every coordinate of ga by \*\*\*\* Lycium chinense.

[Claim 6] The approach characterized by using only the subset of the point about each gaj in an approach according to claim 5 at said step to calculate.

[Claim 7] The approach characterized by using only the correlation sample point set as the object of consideration, and the point which is sharing the edge in an approach according to claim 6 at said step to be used.

[Claim 8] The approach characterized by controlling diffusion of a singular value when calculating a false inverse function further in an approach according to claim 5, in order to calculate the solution to.

[Claim 9] The step which controls said diffusion in an approach according to claim 8 is an approach characterized by sigma1 being the maximum singular value of a matrix in zero, nothing, and here about all less than alphasigma1 singular values.

[Claim 10] The approach characterized by what the 2nd person's face is irradiated with lighting and the structure data which have described the structure of a face, and the reflection factor data which have described the reflection factor property of the face from an exposure are simultaneously incorporated for at the step which prepares the data which define said 2nd face expression set in an approach according to claim 1.

[Claim 11] In an approach according to claim 10 at said step to irradiate Use two or more light sources and one of the light sources of those projects a pattern on the 2nd person's face. At least one of nothing and said the light sources is the infrared light source as a check of the pattern to structure data is possible, and at least one of these the light sources polarizes. At said step to incorporate The approach characterized by using the camera equipped with the polarizer which inhibits specular light and incorporates diffusion component reflective data.

[Claim 12] At the step which prepares the data which define the 2nd face expression set in an approach according to claim 1 The 2nd person's face is irradiated by the 1st polarization light source chosen so that the specular reflection suppression reflection factor property of a face could be checked. The approach characterized by irradiating the 2nd person's face by the 2nd structuring light source which projects a pattern on a face, and incorporating both specular reflection suppression reflective data and structure data from a simultaneous exposure to the exposure and coincidence while irradiating the face by the 1st polarization light source.

[Claim 13] It is the approach characterized by outputting the light of the frequency from which said 1st polarization light source and said 2nd structuring light source differ in an approach according to claim 12.

[Claim 14] It is the approach characterized by said 1st polarization light source and said 2nd structuring light source outputting infrared light in an approach according to claim 12.

[Claim 15] In an approach according to claim 12, the incorporated data are processed further, and it is (a). The data which have described the dimension modality of a face, and (b) Approach characterized by obtaining both data which have described the diffuse reflection property of a face.

[Claim 16] At the step which prepares the data which define said 2nd face expression set in an approach according to claim 1 irradiate the 2nd person's face by the light source from which plurality differs, and range map data are measured from said exposure. The approach characterized by measuring image data from this exposure, drawing a three-dimension side from said range map data, calculating the field normal to said three-dimension side, processing said field normal and said image data, and drawing an albedo map.

[Claim 17] It is the approach characterized by at least one of two or more of the light sources polarizing in an approach according to claim 16.

[Claim 18] It is the approach characterized by two or more light sources of all polarizing in an approach according to claim 16.

[Claim 19] When it performs by computer, the training expression set from a certain person, The expression which corresponds from another person's code book is operated. Said training expression set, The step which calculates the expression and spiral form-conversion corresponding to these, and the step which obtains a synthetic expression set with the application of a transform function from a code book to two or more expression, One or two media or more which are characterized by storing the instruction for performing a computer which can be computer read and which can be computer read.

[Claim 20] It is the medium which is characterized by for said instruction which can be computer read making a synthetic expression set use it for a computer in the medium according to claim 19 which can be computer read, and transforming the expression from one person to the expression of the person of another side and which can be computer read.

[Claim 21] The expression which said instruction which can be computer read is the expression from one person in the medium according to claim 20 which can be computer read, and constitutes the expression of a code book is a medium which is characterized by transforming a different expression to a computer and which can be computer read.

[Claim 22] It is the medium which is characterized by transforming expression to a computer and which can be computer read by transmitting to the receiver with which said instruction which can be computer read can reconstruct expression for at least one index of a synthetic expression in the medium according to claim 20 which can be computer read.

[Claim 23] It is the medium which is characterized by said instruction which can be computer read transforming the expression of a face to a computer in the medium according to claim 20 which can be computer read and which can be computer read.

[Claim 24] The code book which has stored the data which are the code book embodied on [ which can be computer read ] the medium, and define the 1st face expression set of the 1st person, It is data embodied on [ which can be computer read ] the medium, and these data define the 2nd face expression set. This 2nd face expression set Face expression conversion system characterized by having the data used as a training expression set of the 2nd different person from said 1st person, said training expression set, and the conversion processor that draws a transform function from the expression which corresponds from said 1st expression set.

[Claim 25] It is the face expression conversion system characterized by said conversion processor having a linear transform processor in face expression conversion system according to claim 24.

[Claim 26] This synthetic expression set is face expression conversion system characterized by being drawn by applying a transform function to code book expression including the synthetic expression set further embodied in face expression conversion system according to claim 24 on [ which can be computer read ] the medium.

[Claim 27] It is the face expression conversion system characterized by said transform function compensating the size of the face of the 1st person and the 2nd person, and the difference in a configuration in face expression conversion system according to claim 24.

[Claim 28] In face expression conversion system according to claim 24 said conversion processor n expression vector sets of the face which expresses each expression as a 3m vector containing x in the correlation sample location of m pieces, y, and z variation rates, and is changed are ga1... as n The vector set to which a target face corresponds is gb1... When given as n, they are 3m linearity least square system of the following format, and [Equation 2].

$$a_j \cdot g_{ai} = g_{bi}[j], i = 1 \dots n$$

Face expression conversion system characterized by drawing a transform function by \*\*\*\* Lycium chinense by calculating one linear prediction child set aj for every coordinate of ga.

[Claim 29] When the face lighting system which offers the light source from which plurality differs simultaneously since the face of a photographic subject is irradiated, and said face lighting system irradiate, They are the data incorporation system which incorporates both structure data and reflection factor data from the face of a photographic subject, and the 1st composition expression code book. This 1st composition expression code book The training expression set obtained from the photographic subject is received. Said training expression set, The transmitter which has the 1st composition information code book compounded by calculating a transform function using a corresponding non-compounded code book expression, and applying said transform function to all the expression in the 1st composition information code book, The reconstruction module for being the receiver linked possible [ this transmitter and a communication link ], and reconstructing a face image, It has the receiver which has said 1st composition information code book and the 2nd composition information code book which has stored the same synthetic expression. Said transmitter Or it corresponds, in order to incorporate an additional expression of a

photographic subject, and to look for the expression which agrees mostly, the 1st code book is searched. Face expression conversion system characterized by transmitting the index of the code book expression which agrees mostly or it corresponds to said receiver for the face image reconstruction by the reconstruction module.

[Claim 30] It is the face expression conversion system characterized by said lighting system including at least one polarization light source in face expression conversion system according to claim 29.

[Claim 31] It is the face expression conversion system characterized by said lighting system including two or more polarization light sources in face expression conversion system according to claim 29.

[Claim 32] It is the face expression conversion system characterized by including the patternizing light source constituted so that said lighting system might project a pattern on the face of a photographic subject in face expression conversion system according to claim 29.

[Claim 33] It is the face expression conversion system characterized by including the infrared light source to which said lighting system projects a pattern on the face of a photographic subject in face expression conversion system according to claim 29.

[Claim 34] The light source from which said plurality differs in face expression conversion system according to claim 29 is face expression conversion system characterized by all being the infrared light source.

[Claim 35] How to animate the description of the face characterized by being the approach of animating the description of a face, defining the fragmentation side approximated to the geometry of the face from which plurality differs, and fitting the same fragmentation side as each of two or more faces.

[Claim 36] The approach characterized by said step to define defining a fragmentation side with rough mesh structure in an approach according to claim 35.

[Claim 37] It is the approach characterized by rough mesh structure having a triangular mesh in an approach according to claim 36.

[Claim 38] The approach characterized by performing continuation optimization operation covering the top-most-vertices location of a fragmentation side in an approach according to claim 35 at said step to fit.

[Claim 39] The approach characterized by fitting a fragmentation side to a face, without changing the connectability of the mesh which defines the fragmentation side by said step to fit in an approach according to claim 35.

[Claim 40] The approach characterized by making the smoothing functional relevant to the mesh which defines the fragmentation side by said step to fit into the minimum in an approach according to claim 35.

[Claim 41] The approach characterized by choosing one or two constraint or more relevant to the mesh which defines the fragmentation side by said step to fit in an approach according to claim 35, and fitting said constraint directly on the point with which it corresponds on a field.

[Claim 42] It is the approach characterized by relating said constraint with one of an eye, a nose, and the openings in an approach according to claim 41.

[Claim 43] The approach characterized by making the functional which includes the term which expresses distance, smoothness, and constraint with said step to fit in an approach according to claim 35 into the minimum.

[Claim 44] The approach characterized by calculating the solution of a series of linearity least square problems in an approach according to claim 35 at said step to fit.

[Claim 45] One or two media or more which are characterized by storing in one or the two computers or more concerned the instruction which realizes an approach according to claim 35, and which can be computer read when it performs by one or two computers or more and which can be computer read.

[Claim 46] How to animate the description of the face characterized by being the approach of animating the description of a face, defining the fragmentation side approximated to the geometry of the face from which plurality differs, fitting the same fragmentation side as each of two or more of said faces, and changing the expression of one face into the expression of the face of another side for a face and a face using matching and matching with a face and a face.

[Claim 47] Are the approach of animating the description of a face and three-dimension data are measured about the face from which plurality differs. The general-purpose face model used since a corresponding face model is acquired and it maps to each corresponding face model is defined only one. Are directly mapped by the point with which it corresponds on each corresponding face model. Two or more points on said general-purpose face model are chosen, and this general-purpose face model is fitted to each of a corresponding face model. At the step concerned to fit How to animate the description of the face characterized by mapping each of the selected point directly on the point with which it corresponds on each corresponding face model.

[Claim 48] It is the approach which said step to define defined the fragmentation side from base mesh structure, and said fragmentation side has approximation-ized geometry of a face model in an approach according to claim 47. including two or more top-most vertices, and is characterized by operating only the location of the top-most vertices of a fragmentation side at said step to fit.

[Claim 49] The approach characterized by operating the base mesh which defines the fragmentation side by said step to fit in an approach according to claim 47.

[Claim 50] The approach characterized by operating the base mesh which defines the fragmentation side by said step to fit in an approach according to claim 47, without changing the connectability of a base mesh.

[Claim 51] The approach characterized by measuring three-dimension data using a laser range scan in an approach according to claim 47 at said step to measure.

[Translation done.]

\* NOTICES \*

JPO and NCIP1 are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the approach and system of the modeling for animating a face with reality, and a rendering. Specifically, this invention relates to the approach and system for a face image processing.

[0002]

[Description of the Prior Art] In the field of computer graphics, the rendering of various objects is carried out and the object is displayed from a computer display for a user. For example, it is typical that the computer graphics application for generating and carrying out the rendering of the computer object in a computer game, and displaying on a computer screen or television is included. The modeling and rendering of an image with a real feeling serve as an efforts target without an intermission for what is engaged in the computer graphics field. The field which serves as an efforts target especially in the computer graphics field is carrying out the rendering of the face image with a real feeling. As one example, carrying out the rendering of the display of the individual who has participated in conversation is performed with a certain specific computer graphics application. By computer, the image with which the rendering of what is often seen was carried out eventually [ of this individual ] is an image by which the rendering was carried out, and differs from the actual individual often substantially.

[0003]

[Problem(s) to be Solved by the Invention] It is a difficult activity especially by two reasons that carry out the modeling of a face and face expression with a real feeling, and they carry out a rendering. One is because human being's skin has an unsuitable reflection factor property in carrying out modeling with various shading models available now. For example, the well-known Phong model is inadequate for carrying out the modeling of human being's skin. When carrying out the rendering of the face expression and another has a gap also a thing and a little which is sensed as a motion of a "real" face, the gap is because even an indifferent observer is sensed as the not right thing. If the existing motion incorporation (capture) system of a face is used, considerably persuasive face animation can be created, but when the incorporated motion is applied to another face, it has often happened that the persuasive power declines far and becomes very strange. For example, if it takes into consideration that the face which became the origin of the motion is the same face, it is not difficult to animate a motion of each face, if the sampling of a motion of the face is taken from a certain person. Consequently, the description of the motion which is the same between expression other than a certain expression, or is dramatically similar will be acquired. However, it has often happened that it is weak to change a motion of this person's face into another person's face since the difference (for example, the size and the configuration of a face) in a proper is between these two faces. [0004] Therefore, this invention makes it a technical problem to offer the amelioration system and approach for carrying out the modeling of the texture and reflection factor of the skin of human being, and is made. Moreover, by changing a motion of a certain person's face into a motion of the face of an another person object, this invention makes it a technical problem to offer the system and approach for carrying out the reuse of the incorporation (capture) data of a motion of a face, and is made.

[0005]

[Means for Solving the Problem] At the operation gestalt currently illustrated and explained below, it is a three dimension. (3-D) The data which have described the countenance are incorporated, a motion of a face is changed into another individual from a certain individual with reality, and the new technique for carrying out the modeling of the reflection factor of the skin is proposed.

[0006] With the operation gestalt currently explained here, human being's photographic subject is prepared, and since the face of a photographic subject is irradiated, the light source from which plurality differs is used, the structuring light source to which one in two or more light sources projects a pattern on the face of a photographic subject (structured light source) it is . It becomes possible to incorporate the data one or two cameras or more have described the countenance of 3-D of a photographic subject to be according to this structuring light source. Although other one more light source is prepared further, since the face of a photographic subject is irradiated, this is used. The face of a photographic subject is a reflection factor property variously by the other one light sources. (reflectance properties) It checks. Moreover, the light source of this another side is a polarizing filter. (polarizing filter) It is used together and is the specular reflection component of the reflection factor of a face. (specular component) It is removed. Namely, diffuse reflection component (diffuse component) It is incorporated with a camera. If the light source from which plurality differs is used, it will become possible to check simultaneously both the structure property of a face, and a reflection factor property. When the light source is chosen carefully, the light

source is made into a narrowband and it is an adjustment narrowband filter, for example. (matching narrowband filter) If it is used with a camera, it will become possible to remove the effect of the source of perimeter lighting. [0007] Two useful items are obtained as a result of the lighting process currently explained here. Namely, (1) A range map (or depth map (depth map)) and the pattern of (2) structuring light source are the images of the face which is not projected there. Field normal [ as opposed to / 3D side is drawn from a range map and / 3D side ] (surface normal) It is calculated. In order to define 3D side, when processing a range map and the filtering step as an option is included in the processing, it is a general-purpose face template. (genericface template) It will be combined with a range map and an unnecessary noise will be refused. The image of the calculated field normal and a face is an albedo map. (albedo map) It is used in order to draw. an albedo map — special type texture map (texture map) it is — each sample has described the diffuse reflection factor of the face in the specific point on a field (point) there. Therefore, the information which has described 3D side (that is, field normal) of a face at this event of a process, and the information which has described the reflection factor (that is, albedo map) of a face will be checked.

[0008] With a certain operation gestalt, the information or data obtained in the lighting process is used in order to change the expression of a certain person's face into the expression of the face of another person object. At this operation gestalt, it is a code book. (code book) The view is taken in and used.

[0009] Code books are much general-purpose expression of another person object (person A). (generic expressions) The described data are stored. One target is changing the expression of another person object (person B) for the expression of a code book using ejection and its expression. In order to perform this, by the approach by this invention, it trains using Person B. Expression (training expressions) The set is created. Training expression consists of sets of the expression which exists in a code book. It is a transform function when training expression and the code book expression corresponding to each expression are used. (transformation function) It is drawn. This transform function is the set of a synthetic expression which should have agreed on Person's B expression. (a set of synthetic expressions) It is used in order to draw. That is, when a transform function is drawn, this is applied to each of the expression in a code book, and it is made for code book expression to agree on Person's B expression. Therefore, if a synthetic code book expression is searched when a new expression which may not exist in a training set is received from Person B, the expression which agrees most on Person's B expression can be found.

[0010] Common face structure which can use the expression of a face, and a motion with another operation gestalt in order to change into another face from a certain face (common face structure) It defines. the operation gestalt currently explained here — common face structure — rough mesh structure (coarse mesh structure) Or fragmentation side used as a foundation in case it consists of "a base mesh (base mesh)" and this changes a certain person's expression into another person's expression (subdivision surface) The definition is given. A common-base mesh is used to all faces, and 2 or three or more faces are made to be matched. Therefore, if the structure defined by this is used, it will become possible to be adapted for another person object from a certain person in a motion of a face. According to this operation gestalt, the technique for being adapted for the face model of a photographic subject in a fragmentation side is used. By the technique by this invention, some points on a fragmentation side are defined as being directly mapped by the point with which it corresponds on a face model. This is the same also about all different face models that may happen. This constraint (constraint) When it adds, a base mesh will have the property of fitting a different face model similarly. Furthermore, smoothing functional made into the minimum with the algorithm by this invention so that a base mesh and a face model might certainly be matched (smoothing functional) It is used.

[0011] With another operation gestalt, the reflection factor processing technique which metrizes the reflection factor of the face of a photographic subject is prepared. At the technique according to this invention in order to measure a reflection factor, a reflection factor is a diffuse reflection component. (diffuse component) Specular reflection component (specular component) It dissociates and processing of a diffuse reflection component has taken the lead.

[0012] It is an albedo map in order to measure a diffuse reflection component. (albedo map) It defines first. This albedo map is defined by preparing first a camera and the photographic subject irradiated by the light source from which plurality differs. The light source is a polarizing filter. (polarizing filter) It is filtered. This filtering is used together with the polarizing filter with which the polarizing filter was placed in front of the camera, and is performed by preventing that inhibit specular reflection or specular reflection is recorded. An image sequence is photoed so that the head of a photographic subject may be surrounded. Each individual image is processed so that the individual albedo map corresponding to the image may be obtained. All the albedo maps to a specific photographic subject are combined so that a single albedo map may be obtained to the whole face of the photographic subject.

[0013]

[Embodiment of the Invention] Hereafter, the operation gestalt of this invention is explained to a detail with reference to a drawing.

[0014] In order to carry out the rendering of the face expression to a face with an outline real feeling, a very exact model needs to express the reflection factor of the skin, and a motion of a face. Below, the measurement data about geometry, a motion, and a reflection factor is used, and the approach and technique for carrying out the modeling of the face, animating, carrying out a rendering, and reproducing the appearance and face expression of a face of a specific person with reality are explained. Since the perfect model containing geometry and bidirectional reflectance is built, the rendering of the face can be carried out in any lighting conditions and display conditions. Structuring face model with which between different faces was matched by the modeling system explained below

and the approach (structured face model) It is built and this face model serves as a foundation of animation actuation of various faces.

[0015] The operation gestalt by this invention has each part and relation which constitute a face modeling process so that it may explain below. In order to set the geometry expression as the foundation of a face model and to make it become the face modeling of many henceforth, and the foundation of rendering actuation by structuring and unification—izing the expression of geometry, at the operation gestalt of this invention, it is the conventional field adaptation technique. (surface fitting technique) It extends and is a general-purpose face. (generic face) He is trying to make a different individual's face agree. Although the latest rate measurement of a cutaneous reflex is first used in order to build a reflection factor model with a real feeling, this is well-known and is put in practical use.

Furthermore, improved texture incorporation process (texture capture process) It is used and the diffusion texture map measured newly is added. General-purpose mesh (generic mesh) In order to animate, the improved technique is used and the field configuration of having been suitable for the high quality rendering is made to be acquired.

[0016] The outline of the example of the desktop computer 130 which can be used with the operation gestalt currently explained to drawing 1 here as instantiation computer system preliminary knowledge is shown. Under the situation of a distributed computing environment, various things of a computer like a graphic display are usable. These computers carry out the rendering of the graphics according to the explanation indicated below, and they can be used for them in order to process an image (image).

[0017] The computer 130 includes one, two processors or more or the processing unit 132, the system memory 134, and the bus 136 136, i.e., the bus which has combined the various system components containing a system memory 134 with the processor 132. Although the bus 136 is expressed with any one of the several sort type bus structures, or two or more, in this, the memory bus which adopted various bus architectures or a memory controller, the peripheral bus, the high-speed graphics port and the processor, or the local bus is contained. As a system memory 134, it is a read-only memory. (read only memory-ROM) 138 and random access memory (random access memory-RAM) There is 140. Basic input/output system (BASIC input/output system-BIOS) 142 consists of basic routines which support the information transfer between the elements in a computer like [ at the time at the time of a start-up ], and is stored in ROM 138.

[0018] The computer 130 has equipped an optical disk drive 150 still like the hard disk drive 144 written between hard disks (not shown), the magnetic disk drive 146 written between the dismountable disks 148, and CD-ROM and other optical media written between the dismountable optical disks 152. The hard disk drive 144, the magnetic disk drive 146, and the optical disk drive 150 are connected to the bus 136 through the SCSI interface 154 or other kinds of a certain of applicable peripheral interface. These drives and the medium of those which can be related computer read store the instruction which can be computer read, DS, a program module, and other data as non-volatile storage for a computer 130. in the instantiation environment where it is explaining here, although the hard disk, the dismountable magnetic disk 148, and the dismountable optical disk 152 are adopted, if it is this contractor, I will be understood — as — a magnetic cassette, flash memory card, a digital videodisc, and random access memory (RAM) Read-only memory (ROM) etc. — it is also possible to use the medium of other types which can store the data which can access a computer like and which can be computer read in the operating environment of instantiation.

[0019] Some program modules can be stored in a hard disk 144, a magnetic disk 148, an optical disk 152, and ROM138 or RAM140, and operating systems 158 and 1 or two or more application programs 160, other program modules 162, and the program data 164 are contained in it. A user can let input devices, such as a keyboard 166 and a pointing device 168, pass, and can input a command and information into a computer 130. A microphone, a joy stick, a gamepad, a satellite dish, a scanner, one, or two cameras or more are possible as other input devices (not shown). The above and other input devices are connected to the processing unit 132 through the interface 170 combined with the bus 136. The monitor 172 or the display device of other types is also connected to the bus 136 through the interface of a video adapter 174 etc. As for a personal computer, it is typical to have equipped other circumference output devices (not shown) like a loudspeaker or a printer other than a monitor.

[0020] Usually it operates by the networking environment which uses a logic connection with one or two remote computers [ like a remote computer 176 ] or more whose computer 130 is. another personal computer as a remote computer 176, a server, a router, Network PC, and a pier — although it is possible to make it a device or other common network nodes and only the memory storage device 178 is shown in drawing 1 , it is typical for there to be many elements mentioned above in relation to the computer 130, or to have equipped all. As a logic connection who shows drawing 1 , it is a Local Area Network. (local area network-LAN) 180 and wide area network (wide area network-WAN) There is 182. This kind of networking environment is office, the computer network in an enterprise, intranet, and the Internet. (the Internet) It is the then common existence.

[0021] When used by LAN networking environment, the computer 130 is connected to the local network 180 through the network interface or the adapter 184. When used by WAN networking environment, in order to establish communication on a wide area network 182 like the Internet, as for a computer 130, it is typical to have equipped other means like a modem 186 or a network interface. A modem 186 has built-in and an external mold, and is connected to the bus 136 by both through the serial port interface 156. At networking environment, the program module mentioned above in relation to the personal computer 130 or its part can be put on a remote memory storage device. The network connection of a graphic display is instantiation and it is also possible to use other means to establish the communication link between computers so that I may naturally be understood.

[0022] Generally, the data processor of a computer 130 is programmed with the instruction of a computer variously



stored in the medium which can be computer read by differing in time amount, respectively. As for a program and an operating system, being decentralized on a floppy disk or CD-ROM is typical. A program and an operating system are installed or loaded to the secondary memory of a computer from a floppy disk or CD-ROM. At the time of activation, as for a program and an operating system, the part is loaded to the primary electro nick memory of a computer at least, the type which mentioned above the instruction or program for realizing the step currently explained below in relation to a microprocessor or other data processors according to this invention currently explained here, or others — many things are stored in the storage of a type which can be computer read. Moreover, according to this invention, the computer itself is programmed according to the approach and technique of being explained below.

[0023] Like the expedient top of explanation, and an operating system, although a program and other executable-program components are shown as a discrete block here, these programs and components differ in a stage, are put on the different-species storage component of a computer, and are performed by the data processor of a computer so that I may naturally be understood.

[0024] Conventionally [ instantiation system ] for incorporating the structure and the property of the face, it is an incorporation system. (capturing system) Both the face structure of the whole face and a reflection factor property were not able to be independently incorporated simultaneously from mutual. For example, structuring light (structured light) These systems are reflection factors although the system which uses it and incorporates the structure of a face exists. (reflectance) The property [ like ] of a face is not incorporated. As for these systems, the structure of a face is not incorporated although similarly the system which incorporates the reflection factor of a face exists. If the structure and the reflection factor of a face can be incorporated simultaneously independently mutually, another actuation will be added to collected data and it will become possible to use for various face renderings and animation actuation. Although one example of instantiation rendering actuation is explained, the information or data obtained as a result of the system and approach of lower-\*\*(ing) can be used in other various fields below so that I may naturally be understood. For example, construction of the face model which has a real feeling aiming at animation, such as recognition of the face aiming at security, the exchange between personal users, etc., a game, and a movie (a film, animation), as an application field, and a user incorporate their face easily, and are made to be able to use it by interactive entertainment or business communication.

[0025] Drawing 2 shows the instantiation system 200 which is suitable for using it when incorporating simultaneously the face structure and the reflection factor property of a face of a photographic subject. This instantiation system equips the data incorporation system which carried out the form of one or two cameras or more, and one of them is illustrated as a camera 202. A camera 202 equips a CCD image sensor and a related circuit, operates a CCD array, reads an image in a CCD array, changes an image into a digital form, and makes it possible to notify the image to a computer. Moreover, a system is a face lighting system which carried out two or more light sources or the form of a projector. (facial illumination system) It has equipped. It is a JIEN lock so that these cameras may make simultaneous incorporation possible in time, when two or more cameras are used. (genlock) It is carried out. The two light sources 204 and 206 are used in the example of a graphic display. The light source 204 is a structuring pattern (structured pattern) projected on the face of a photographic subject desirably. It outputs. If the light source 204 is a suitable location, it is possible to place anywhere. This pattern makes it possible to incorporate the structure information or data about 3D configuration of a face of a photographic subject with a camera 202. If the pattern which consisted of light of an infrared field is adopted, it is advantageous, but if it is the suitable light source, it is usable at every light source. The light source 206 outputs desirably the light which enables it to incorporate the diffuse reflection component of the reflection factor property of a face with a camera 202. The light source 206 is for example, a beam split. (beam splitting) If it puts on lens 202a of a camera, and the same rank by technique, it is advantageous, but placing anywhere is possible if it is a suitable location. As for this light source, it is possible to also make it adapted so that the lens of a camera may be surrounded. Moreover, this light source is the specular reflection component of a reflection factor. (specular component) It is chosen so that it may be inhibited or removed. At the example of a graphic display, it is a linearity polarizing filter. (linear polarizing filter) Since it is adopted in order to obtain polarization lighting, and another polarizing filter has sense which intersects perpendicularly with said linearity polarizing filter and it is placed before lens 202a, the specular reflection from a face is made not to be recorded with a camera. The lighting system mentioned above is simulated using the light source of a different frequency, for example, the light source corresponding to the red and the green channel of a camera. In addition, both of these channels can be made an infrared field. Furthermore, if it chooses so that the light source may become a narrowband (for example, 780-880 nm), the effect of an ambient light is removable. It is only a time of being filtered so that a camera may also become a narrowband that this property is attained. Since an ambient light is diffused to the lighting from the light source being concentrated on the wavelength of a narrowband covering the wavelength of a broadcloth range, the light from the light source becomes more powerful than the ambient light at the time of the above wavelength. Since a camera is filtered so that only the wavelength emitted from the light source may be recorded, being influenced by the ambient light does not almost have it. Consequently, a camera will detect only the effect the selected light source affects a photographic subject.

[0026] If the light source from which plurality differs is used, if the infrared light source is used together with the polarization light source (this can also be made the infrared light source), specifically, it will enable a camera to incorporate independently the structure information about a face, the reflection factor information about data and a face (from the light source 204), or data (from the light source 206) simultaneously mutually. Reflection factor information has described the diffuse reflection factor property of a face to structure information having described

the three-dimension modality of a face. When this information is processed by computerization image processors, such as a computer 208, the information or data which can be used by future face animation actuation will be obtained. At the example explained below, this information is three-dimension data (3D data) and an albedo map. (albedo map) It contains.

[0027] Drawing 3 is a flow chart which has indicated the step of the approach by the operation gestalt explained here. According to the approach explained here, the information or data about the structure and the reflection factor property of a face are collected simultaneously, and it is possible to process. It irradiates by the light source from which, as for the face of a photographic subject, plurality differs at step 300. The instantiation system which realizes this step is shown in drawing 2. In addition, although the two instantiation light sources are used in the example of a graphic display, it is also possible to use two or more of other light sources so that I may naturally be understood. At step 302, range map data (depth map data) and image data are measured from the exposure of step 300. That is, it will be used, in order that a camera may be enabled to detect the rate of a light reflex and it may obtain both image data (that is, reflection factor) in which this rate of a light reflex does not contain the structure light source pattern range map data and there by the exposure of step 300. Range map data and image data are passed to a computer 208 ( drawing 2 ), and are processed there. At this event, it is step 304, and is a general-purpose face template. (generic face template) This is an option, although the various noises produced in relation to range map data are removable if it applies to range map data. A general-purpose face template can be considered to be 3D filter from which the noise contained in range map data is removed. A general-purpose face template is obvious to the familiarity person of this field.

[0028] At step 306, range map data are used, and 3D side is drawn or calculated. If it is a suitable algorithm, it is usable with every algorithm and this is obvious to the familiarity person of this field. There are some which are indicated by the following paper as an example of an algorithm. namely, Turk & Levoy work "zipper DOPORI gon mesh (Zippered Polygon Meshes) from a range image" (SIGGRAPH94) and F.Bernardini, J.Mittleman, H.Rushmeier, C.Silva, and G.Taubin collaboration "baud ruby BOTTINGU algorithm (The Ball-Pivoting Algorithm) for field reconstruction" (Trans.Vis.Comp.Graph.5:4 (1999)) it is . At step 308, the field normal vector ("field normal") to 3D side of step 306 is calculated using a well-known algorithm. One method of performing this task is calculating the normal to a triangle, taking the average of the triangle normal of the circumference of each top-most vertices, making a top-most-vertices normal, and interpolating a top-most-vertices normal so that the interior of three square shapes each may be crossed. Although it is natural, it is also possible to use other approaches. Next, at step 310, the image data of the field normal calculated at step 308 and step 302 is used, and an albedo map is drawn. An albedo is a special type texture map and each sample has described the diffuse reflection factor of the face in the specific point on the face there. Derivation of an albedo map is obvious to the familiarity person of this field, if the information shown above is given. The algorithm of instantiation is Marschner work "reverse rendering (Inverse Rendering for Computer Graphics) of computer graphics" (PhD thesis, Cornell University, and August 1998). It is indicated.

[0029] At this event, as shown in drawing 2, 3D data which have described the structure description of the face of a photographic subject, and the albedo map data which have described the diffuse reflectance of a screen are obtained as a result of lighting processing.

[0030] Since the above-mentioned lighting processing can be used in order to extract the information described, the extracted information can be used for the suitable object. With the operation gestalt with especially an advantage, the extracted information is used in order to extract and recognize the expression of a photographic subject. This information can be used for expression conversion. With the operation gestalt by this invention currently explained immediately in the back, a certain person's expression can use the expression of an another person object in order to have and change a real feeling.

[0031] At the expression conversion \*\*\*\* expression conversion implementation gestalt by the activity of a code book, it is a code book. (code book) A view is taken in and it is used by expression conversion actuation in which it is explained below. Drawing 4 shows the example of the code book 400 which has stored a different expression of a large number incorporated by a certain person. Such expression is general-purpose expression. (generic expressions) That is, the expression from a general-purpose person can be considered from a specific individual. In this example, expression has the range from expression 1 to Expression N. Expression 1 is a "smile", expression 2 is at a "wry look" and expression 3 is the expression of the "resentment." Hereafter, it is the same. A thing like the process which it is described in the form of those geometry and explained mathematically immediately in the top to incorporate by the suitable approach is possible for the expression stored in the code book 400.

[0032] In order to perform expression conversion, a part of expression in a code book 400 is used, and it is a transform function. (transformation function) It is drawn first. In order to draw a transform function, it is a training expression set. (training set of expressions) The view 402 is taken in. The expression training set 402 consists of sets of the expression obtained from an individual with the expression other than the individual described by the code book 400. The training expression of the training set 402 is the subset (subset) of code book expression. That is, each expression in a training set supports the expression in a code book 400. For example, the training set 402 can be constituted from three expression 1, i.e., expression, expression 2, and expression 3, and expression has become a "smile", a "wry look", and the "resentment" there, respectively. The target of a transform function applies the geometric deformation related with the expression of a training set to reception, applies the geometric deformation to all the expression of a code book 400, and is making it code book expression serve as an expression expression with a real feeling. That is, it is thought that the face of an everybody object carries out the method of

deformation that geometry differed, also with the expression of which arbitration. When deformation of a certain person's geometric face at the time of a certain expression is simply applied to the face of another person object in order to carry out the rendering of the expression, the face to which the deformation was applied will seem to have been distorted dramatically. The geometry of a face not only differs, but such a thing happens, because deformation of a face differs like [ between faces ]. Therefore, a transform function with which the best conversion to another expression set from a certain expression set is obtained is drawn.

[0033] If it returns and explains to drawing 4, the linear transform processor 406 is shown in drawing 4. This conversion processor 406 can be realized in suitable hardware, software, firmware, or such combination. The linear transform processor 406 is realized by software in the example of a graphic display. A linear transform processor receives the training expression set 402 and the corresponding code book expression 404 as an input. A conversion processor processes these inputs and draws a transform function 408. When this transform function 408 is applied to all the expression in a code book 400, it is a synthetic expression set. (synthesized set of expressions) It is obtained. This synthetic expression set expresses the expression of a code book operated according to the geometric deformation relevant to the expression of the person by whom the training expression set was got.

[0034] The variation rate of a face when expression is the same is not the same in a person who is different by the following two reasons. It is because one of them is not in agreement with accuracy if the motion incorporation sample point (it is shown by this example according to this specific algorithm how a motion of a face can express) has an error in a location. Another is because the configuration and size of the head change between persons.

[0035] The 1st inequality is cancelable by all faces' moving, incorporating and re-sampling displacement data by the fixed location set on a general-purpose mesh. This is explained to the part of the following "instantiation systems for building a face model and approaches" in detail. There, the fixed location set is called "the correlation sample location (standardsample positions)." a re-sampling function — mesh deformation function (mesh deformation function) it is. Correlation sample locations are top-most vertices of the face mesh corresponding to top-most vertices of the general-purpose mesh subdivided once.

[0036] It is necessary to change displacement data into another face from a certain face, and to compensate the size of a face, and change of a configuration with the 2nd inequality. In the example of a graphic display, this is performed by finding the best linear transform from one side to another side, after finding the small training set of the expression which corresponds about two data sets. The following case is considered as one example. That is, in the experimental environment, the label was manually attached to 49 corresponding expression in which the expression showing feeling contains various strength of two or more expression. About a motion of voice, 10,000 frames is time amount distortion. (time warping) Alignment was automatically carried out using technique.

[0037] Each expression is expressed by the 3m vector  $g$  containing  $x$  in the correlation sample location of  $m$  pieces,  $y$ , and all the  $z$  variation rates. They are [ Every one set  $a_j$  is calculated for every coordinate of  $g_n$  by calculating the solution of 3m linearity least square system shown below. ]  $n$  expression vector sets of the face changed  $ga_1 \dots$  It is the vector set to which it is referred to as  $n$  and a target face corresponds  $gb_1 \dots$  When  $n$ , he is a linear prediction child. (linear predictors)

[0038]

[Equation 3]

$$a_j \cdot g_{ai} = g_{bi} [j], i = 1 \dots n$$

[0039] In the example of a graphic display, only the subset with each small point of  $g_{aj}$  is used. Specifically, only the correlation sample point and the point which shares an edge are set as the object of consideration. In the used mesh, it is \*\*\*\*\* (average valence) Since it is about 6, typical  $g_{aj}$  is 18 elements. The system of the result is  $n \times 18$  generally.

[0040] The linear system of a result may not have enough conditioning and the linear prediction child  $a_j$  is not conceptualized good in that case. Singular value (singular value) Diffusion is a false inverse function. (pseudoinverse) Since it is controlled when calculating and calculating the solution to, conceptualization is improved substantially. All less than  $\alpha$  singular values are made into zero.  $\sigma_1$  [ however, ] — the maximum singular value of a matrix — it is —  $\alpha = 0.2$  and ... it is  $0.1$ .

[0041] Drawing 5 is a flow chart which has indicated the step of the expression conversion approach by the operation gestalt currently explained here. The code book of expression is prepared at step 500. The example of this kind of code book is shown above. A training expression set is prepared at step 502. As an example of representation, this training set is the set of the expression from a different person from the person by whom code book expression was got. If the training expression set is appropriate, it can be incorporated by every approach. As one example, expression can be incorporated using a system as shown in drawing 2. After a training expression set is prepared, a transform function is drawn at step 504 using a training set and a code book. One example of an approach which does this activity is as having mentioned above. Although it is natural, unless it deviates from the pneumonia and the range of this invention, it is also possible to use other approaches, for example, neural network (neural network) etc. — also using various nonlinear conversion — basic expression (basis expressions) It is also possible to use a weighted sum. If a transform function is drawn, this is applied to all the expression in a code book, and the synthetic expression set which can be used as a foundation of future face animation actuation will be obtained, or it will be defined.

[0042] Application drawing 6 of instantiation shows a system 600. There, it is shown by one example how the expression translation process mentioned above is adopted. a system 600 — the transmitting-side computing

system 602, i.e., a transmitter, and the receiving-side computing system 604, i.e., a receiver, — containing — these — the Internet (the Internet) etc. — it connects so that it may communicate through a network 603. It is constituted so that a transmitter 602 may incorporate a certain person's expression, as this was mentioned above with reference to drawing 2 including the lighting system 200 ( drawing 2 ). Furthermore, the transmitter 602 contains the same code book 400 as what was mentioned above with reference to drawing 4 . In addition, the code book shall be compounded by the synthetic expression set here, as mentioned above. That is, the training expression set obtained from the person with the expression which the lighting system 200 incorporated is used, a code book is processed, and the synthetic expression set is obtained.

[0043] It is constituted so that a face image may be reconstructed from the data with which the receiver 604 received this from the transmitter 602 including the reconstruction module 606. Moreover, the receiver 604 contains the same code book 400 as it by the side of a transmitter 602. Here, the person who is in a transmitter 602 side is going to communicate with the person who is in a receiver 604 side. If a face is moved for a communication link of the person who is in a transmitter 602 side, the face expression and motion will be incorporated and processed by the transmitter 602. It is possible to incorporate face expression in this processing, to search a synthetic code book to it, and to include to find the agreement expression most approximated in a code book in it. If the agreeing expression is found in a synthetic code book, the index of the expression will be sent to a receiver 604, and the animated face will be reconstructed using the reconstruction module 606.

[0044] Face conversion drawing 7 of instantiation is drawing showing some effectiveness of the expression conversion by the operation gestalt currently explained here. The picture (A) of the 1st line constitutes the 1st person's (person A) synthetic face, and three different expression is shown there. These pictures are as a result of the motion of a face by which Person A was incorporated. The motion of the 2nd person's (person B) face is already incorporated. The motion of a face by which Person B was incorporated is shown in the 3rd line (C). Here, 3D motion data are incorporated by measuring a motion of a dot, when some color dots are placed on a person's face and a person's face is deformed, but naturally this will be understood, if this field becomes a familiarity person. In addition, a motion of a face can be incorporated by the system and approach which were mentioned above. The motion by which Person B was incorporated is used as mentioned above, and Person's A expression is changed. The result of this actuation is shown in the 2nd line. All the expression of a these three sets picture is mutually in agreement. As shown in this drawing, even if the expression of the 1st line (A) and the 2nd line (B) is drawn from two dramatically different persons, on the whole, the expression of the 1st line (A) and the 2nd line (B) is not similar for the 2nd person's (the 3rd line (C)) original copy expression to being dramatically similar.

[0045] The model of the face needed in order to make the instantiation system for building a face model and an image with an approach real feeling consists of two parts. The amount of [ of a model ] part I is related to the geometry (that is, configuration of the face) of a face. On the other hand, the amount of [ of a model ] part II is related to the reflection factor (that is, the color and reflection property of a face) of a face. This section explains a part for part I of this model, i.e., the geometry of a face.

[0046] The geometry of a face consists of what added another field which expresses an eye with a hide skin surface. A hide skin surface is drawn by scanning the head in a laser range, and is expressed with this example in the fragmentation side with a displacement map. An eye is another model, alignment of this is carried out to a hide skin surface, it is merged, and the perfect face model suitable for a high quality rendering is obtained.

[0047] At the 1st step which builds a mesh adaptation face model, the fragmentation side closely approximated to the geometry measured with the laser range scanner is created. The fragmentation side is defined from a rough triangular mesh by the example of a graphic display using the fragmentation rule of Loop. The fragmentation rule of Loop is Charles Loop work "smooth fragmentation side (Smooth Subdivision SurfacesBased on Triangles) based on a triangle" (PhD thesis, University of Utah, and August 1987). It is explained in detail. Furthermore, a fragmentation side is work "partition smooth side reconstruction (Piecewise smooth surface reconstruction)" (Computer Graphics (SIGGRAPH'94 Proceedings) pp.295-302 and July 1994) besides Hoppe. The Sharp edge of what is indicated, and resemblance is added and it defines. In addition, irregular crease mask (crease mask) It is not used. Furthermore, darts (dart) When the edge between crease top-most vertices (crease vertex) is subdivided, only the new edge which adjoins crease top-most vertices is marked as a Sharp edge.

[0048] The fragmentation side over all face models is defined using a single base mesh, and it is changing so that only a top-most-vertices location may be adapted for the configuration of each different face. In the example of a graphic display, the base mesh with 227 top-most vertices and the triangle of 416 pieces has the whole face configuration, it is defined as detailing an eye and near a lip, and the most complicated geometry and the most complicated motion have appeared there. The aperture of opening is the boundary of a mesh, and the adaptation process period remains closing by connecting the location of the top-most vertices where it corresponds on the upper lip and a metastoma to one. There is an edge of the fraction marked on sharp fragmentation rules in this base mesh, these create a corner on both sides of the aperture of opening, and it is useful to obtaining the location when turning up the both sides of a nose. Since the crease is taken in only to the chain of at least three Sharp edges with the corrected fragmentation rule, this model does not have a crease in a front face. Isolated top-most vertices (isolated vertex) Limit normal defined clearly (limit normal) It is because it cannot have.

[0049] Drawing 8 shows the example of the rough definition mesh (central drawing) used according to this example. If drawing 8 uses this rough mesh how and maps the same fragmentation control (\*\*) mesh in the displacement fragmentation side of each face, it shows whether natural response relation is obtained between a certain face and another face. This side face is explained in detail below.

[0050] The process used in order to fit a fragmentation side to each face is based on the algorithm indicated by work "partition smooth side reconstruction (Piecewise smooth surface reconstruction)" (Computer Graphics (SIGGRAPH'94 Proceedings) pp.295-302 and July 1994) besides Hoppe. Fundamentally, the field suiting method of Hoppe can be said to be what consists of three phases. That is, they are topology mold presumption (phase 1), mesh optimization (phase 2), and partition smooth side optimization (phase 3).

[0051] In a phase 1, the triangular mesh in which the unorganized point set is given to a strange field top or the neighborhood and which consists of many triangles relatively is built. In this phase, the topology mold of a field is judged and initial presumption of geometry is obtained. A phase 2 begins from the output of a phase 1, reduces the number of triangular, and improves the adaptation to data. In this approach, it is making to optimize the energy function which models clearly a trade-off between the targets to compete mutually called expressional compaction and good-izing of adaptation into the technical problem. As a free variable in this optimization procedure, they are the number of the top-most vertices in a mesh, and its connectability. (connectivity) And that location exists. A phase 3 fits an exact and brief partition smooth fragmentation side by leaving the optimization mesh (partition linearity side) obtained in the phase 2, and optimizing the energy function which trades off simplicity and the compatibility to data also in this case. If optimization is performed in a phase 3, the number of the top-most vertices in a control mesh, its connectability, its location, and the number and location of the sharp description will change. The sharp automatic detection of the description and the recovery in a field serve as an indispensable part of this phase.

[0052] In this operation gestalt, the processing method differs from the approach indicated by Hoppe etc. at two points. Since the 1st does not need to change the connectability of a control mesh, it is that continuous optimization is performed only to a top-most-vertices location. Furthermore, constraint of the description is added with the smoothing term.

[0053] In the example of a graphic display, an adaptation process minimizes the following functional.

[0054]

[Equation 4]

$$E(\mathbf{v}) = E_d(\mathbf{v}, \mathbf{p}) + \lambda E_s(\mathbf{v}) + \mu E_c(\mathbf{v})$$

[0055] In the above,  $\mathbf{v}$  is the vector of all top-most-vertices locations, and  $\mathbf{p}$  is the vector of all the data points from a range scanner. The inferior letter of the 3rd term expresses distance, a configuration, and constraint. A distance function  $E_d$  is the sum square distance from the range scanner point to a fragmentation side. (sum-squared distance) It measures.

[0056]

[Equation 5]

$$E_d(\mathbf{v}, \mathbf{p}) = \sum_{i=1}^{n_p} a_i \|p_i - \Pi(\mathbf{v}, p_i)\|^2$$

[0057] Setting above,  $p_i$  is the  $i$ th range point and  $\Pi(\mathbf{v}, p_i)$ . It is projection of the range point on the fragmentation side defined by the top-most-vertices location  $\mathbf{v}$ . For weight  $a_i$ , the eye direction of the scanner in  $p_i$  is  $s(p_i)$ . It is the Boolean term which makes the point which is not in agreement with the field normal which can be set disregarded. Furthermore, the point which is distant from the field beyond a certain distance is refused.

[0058]

[Equation 6]

$$a_i = \begin{cases} 1 & \text{if } \langle s(p_i), n(\Pi(\mathbf{v}, p_i)) \rangle > 0 \text{ and } \|p_i - \Pi(\mathbf{v}, p_i)\| < d_0 \\ 0 & \text{otherwise} \end{cases}$$

[0059] It sets above and is  $s(p)$ . The direction and  $n(x)$  which go to the view of the scanner in Point  $p$  It is the field normal which faces the outside in Point  $x$ .

[0060] The smooth function  $E_s$  carries out flattening of the control mesh locally. This function measures the distance to the average of each top-most vertices to adjoining top-most vertices.

[0061]

[Equation 7]

$$E_s(\mathbf{v}) = \sum_{j=1}^{n_v} \left\| v_j - \frac{1}{\deg(v_j)} \sum_{i=1}^{\deg(v_j)} v_{b_i} \right\|^2$$

[0062] Top-most vertices  $v_{k_i}$  are the adjoining top-most vertices of  $v_j$ .

[0063] A constraint function  $E_c$  is the sum square distance to the target location set which corresponds from a constraint top-most-vertices set simply.

[0064]

[Equation 8]

$$E_c(\mathbf{v}) = \sum_{i=1}^n \|A_i \mathbf{v} - \mathbf{d}_i\|^2$$

[0065] In the above,  $A_j$  catches and defines the marginal location of the  $j$ -th top-most vertices from a control mesh, and the marginal location of top-most vertices  $c_i$  is the linear function added to the 3D point  $d_i$ . Although it is also possible to apply constraint strictly by re-parameterization of an optimization variable, if this constraint is used as software, it will have become clear that a repeat is smoothly performed to the desirable local minimum. Constraint is chosen by the user in order to make the face description of a general-purpose mesh agree with the description to which it corresponds on the suiting specific face. In this example, constraint of about 25 thru/or 30 is used and it is concentrating on an eye, a nose, and opening. Drawing 9 indicates the point with which it corresponds on a face model (B) to be constraint of the fragmentation control mesh (A) of 900.

[0066]  $E(\mathbf{v})$  Since  $p_i$  and  $a_i$  are not the linear functions of  $\mathbf{v}$ , it is a nonlinear least square problem to make it the minimum. However,  $a_i$  is kept constant and it is  $p_i(\mathbf{v}, p_i)$ . This can be made into a linearity problem if it approximation-izes with the combination of fixed linearity-control top-most vertices. Therefore, an adaptation process will be advanced as a sequence of a linearity least square problem, and projection of  $p_i$  on  $a_1$  and a field will be re-calculated in front of each repeat. A fragmentation marginal side is approximation-ized by the mesh in specific fragmentation level in order to perform these count. Suiting a face is performed by a small number of repeat (less than 30), and with progress of a repeat, constraint is updated according to a simple schedule and goes. The updating begins from high  $[\lambda]$  and low  $[\mu]$ , will lead optimization to approximation-ization of a very smooth face, will go, and will progress to low  $[\lambda]$  and high  $[\mu]$ , and data and constraint will suit closely by the final solution. Actual computation time is decided by computation time of  $p_i(\mathbf{v}, p_i)$ .

[0067] In order to make the mesh for a rendering, a field is subdivided to desirable level and the mesh smoothly approximated to the configuration of a face is obtained. A variation rate is calculated by making the line normal to the field in the top-most vertices intersect the triangular side defined by the original copy scan for every top-most vertices. This is explained to work "displacement fragmentation side (Displaced Subdivision Surfaces)" (SIGGRAPH'00 Proceedings and July 2000) besides Lee. Although the field of the result is reproducing all the descriptions with which the original copy scan was conspicuous in the mesh whose triangle has decreased a little, this is because the triangle of base mesh of the field where a face is important has increased. Inclusion multiplex resolving basis function set which was defined as that which parameterized the field by that parameterization according to the expression of this fragmentation base (built-in set of multiresolution basis functions) It will be obtained, and further, since the description constraint is used at the time of adaptation, natural response relation will be obtained over all the faces that suit using this approach. If you use this structure by the animation of a face, it is convenient in many respects.

[0068] Drawing 10 is a flow chart which has indicated the step of the approach for building a face model according to the operation gestalt currently explained here. This approach can be realized in suitable hardware, software, firmware, or such combination. The approach is realized by software in this example.

[0069] At step 1000, 3D data of one or two faces or more are measured, and a corresponding face model is obtained. In the above-mentioned example, 3D data were obtained by scanning a face in a laser range. 3D data can also be obtained by other approaches so that I may naturally be understood. The general-purpose face model used since one or two face models or more are suited is defined by step 1002. A general-purpose face model has the advantage that it can use so that the face from which a large number differ may be suited so that I may naturally be understood. Therefore, this serves as an improving point compared with the conventional approach that such a thing was not performed. At the example mentioned above, a general-purpose face model is a rough triangular mesh. (coarse triangle mesh) It has mesh structure which carried out the form. This triangular mesh defines the fragmentation side closely approximated to the geometry of a face. The fragmentation side about all the face models is defined by the example of a graphic display using a single base mesh. At step 1004, each point or constraint on a general-purpose face model is chosen. These points or constraint are directly mapped by the corresponding point marked on the face model. [each] Mapping of the point of these each is similarly performed about each of a different face model of a large number which may happen. At step 1006, a general-purpose face model suits one or two face models or more. This step is performed by operating only the location of top-most vertices so that it may be adapted for the configuration of each different face. Since continuous optimization is performed only covering a top-most-vertices location, the continuity of a mesh is not changed into an adaptation process period. Furthermore, in this adaptation process, each point or constraint is directly mapped by the face model. Furthermore, smooth term (smoothing term) It is added, and this is minimized so that flattening of the control mesh may be carried out locally.

[0070] The displacement fragmentation side as for which the eye has carried out additional \*\*\*\* expresses the configuration of the hide skin surface of a face very well. However, there are some other descriptions made desirable in order to obtain a face with a real feeling. The most important description in it is an eye. Since a laser range scanner does not incorporate the suitable information about an eye, a mesh is reinforced by adding the eye modeled independently for the rendering. Since an eye and its motion are not measured from a specific person unlike other parts of a face model, the appearance of a real eye is not necessarily reproduced. However, existence of an eye and a motion are dramatically important for the overall appearance of a face model.

[0071] If it is a suitable eye model, an eye can be modeled using the eye model. In the example of a graphic display, in order to build the model which consists of two parts, the commercial modeling package is used. The amount of part I is the model of an eyeball, the amount of part II is the model of the hide skin surface of the circumference of

an eye, and a part for a palpebra, an eye socket, and perimeter regions of face is contained in it (the amount of this part II will call it "the orbital surface (orbit surface)"). In order to use an eye as some whole face models, it is necessary to fit the orbital surface to the individual face which it is going to model, and to sew up two fields. This is performed at two steps. First, distortion is attached according to the weighting function with which two mesh was defined on the orbital surface, and it is made to appear simultaneously in the part where the eye socket overlapped the face. Next, 2 sides are omitted by this cardiac ellipsoid of a pair, and it joins by stitching and they are carried out so that it may become one mesh.

[0072] In addition, according to these operation gestalten, one of the descriptions with the advantage of the operation gestalt mentioned above is that the structure or the framework which can be used in order to change a certain person's expression into another person's expression is acquired. Suiting an individual's face in a general-purpose face model is restrained so that every part of a general-purpose model may always be mapped by the same description on the face of an everybody object. For example, the left corner of opening in a general-purpose model is always mapped by the left corner of the opening by every person's face. If the set of an adaptation face model is used, in every face, the point corresponding to the specific point in a certain specific face can be judged. For example, suppose that the motion of the left corner of opening on Person's A face was measured. If it can judge which point of a general-purpose model is equivalent to the measure point if adaptation of the general-purpose model to Face A is used, next adaptation of the general-purpose model to Face B is used, it can judge which point of the face of B is equivalent to the point with which it was calculated on the general-purpose model, therefore whether the measure point of Face A is supported. Since it is necessary to know it is necessary which part of a face new in order to reproduce the point set lost motion on the measured face to move, this information is indispensable when changing a motion into another face from a certain face.

[0073] A motion of the motion face of a face is specified by 3D location which changes with the time amount of the sample point set on a screen. When a face moves and it is controlled by incorporation data, these points are the markers on a face by whom tracking is done by the motion incorporation system. The movement toward these points is used in order to control the face by the set of a control point which influences the field of the face smoothly. Incorporation of the motion data of a face can be performed by one of suitable approaches so that I may be understood, if this field becomes a familiarity person. At one example, a motion of a face is work "creation (Making Faces) of a face" (Proceedings of SIGGRAPH 1998, pp.55-67, 1998) besides Guenter. It was incorporated using the technique indicated.

[0074] The mesh deformation face is animated according to the linearity-combination of the variation rate of the control point set  $q_j$  by displacing each top-most vertices  $w_i$  of a triangular mesh from the static position. These control points correspond by the sample point  $p_j$  which has described the motion, and 1 to 1. When the effect each control point affects top-most vertices falls with the distance from the corresponding sample point and two or more control points influence top-most vertices, those weight is normalized so that total may be set to 1.

[0075]

[Equation 9]

$$\Delta w_i = \frac{1}{\beta_i} \sum_j \alpha_{ij} \Delta q_j \quad ; \alpha_{ij} = h(\|w_i - p_j\|/r)$$

[0076] the case where top-most vertices  $i$  are influenced by two or more control points in the above —  $\beta_{i,j} = \sum_k \alpha_{ik}$  it is — otherwise, it is 1. Since such weight is calculated once using the static position of the sample point and a face mesh, a motion of the mesh for every frame is sparse-matrix multiplication. (sparse matrix multiplication) It has become. The degree type was used about the weighting function.

[0077]

[Equation 10]

$$h(x) = 1/2 + 1/2 \cos(\pi x).$$

[0078] In order to process the details which animate a face, the 2 type exception is prepared in these weighting rules. The control point which the tag of a "top" and the "bottom" is attached, for example, the neighboring top-most vertices and the neighboring control point of an eye and opening have on opening does not influence a motion of the top-most vertices under opening. Moreover, the Scala texture map in the field around an eye (scalar texture map) It is used in order to carry out weighting to a motion so that an eye may become the shape of a smooth taper and may become zero in the part of a palpebra. In order to move a face mesh according to a sample point set, a control point location is calculated and, as it is, the face must be made to deform. If the same weighting function as having mentioned above is used, it is calculable what kind of motion the sample point carries out according to a control point. The result is linear transform  $p = Aq$ . Therefore, in order to ask for the sample location  $p_t$  at time of day  $t$ , control point  $q_t = A^{-1}p_t$  can be used. However, since Matrix  $A$  has that conditioning is not enough, in order to avoid the field configuration which happens since the motion of a control point is very large and which is not desirable,  $A^{-1}$  is calculated using SVD (Singular Value Decomposition-singular value decomposition), and the singular value of  $A^{-1}$  is clamped to Limitation  $M$ .  $M = 1.5$  is used in the example of a graphic display, as the reference which is discussing SVD — Golub and Van Loan work "matrix count (Matrix Computations)" — there are the 3rd edition, Johns Hopkins press, and 1996.

[0079] In order to wear the motion face of an eye and the head the appearance near thing, the motion generated by the procedure is added to an eye, and the rigid body motion incorporated independently is added to the whole head.



An eyeball rotates according to the sequence of the random fixed direction, from one eyeball, toward the following eyeball, moves smoothly and goes. A palpebra rotates the top-most vertices which define the palpebra as the circumference of the shaft passing through the core of an eyeball, and is animated because deformation is made to be performed smoothly using the weight defined on a palpebra mesh.

[0080] The rigid body motion of the head is incorporated from a physical motion of a person's head. This While the person is wearing the hat marked with the special machine cognitive possible target, it is carried out by copying the motion on a film (a hat). Work besides Marschner "the image base BRDF measurement (Image-based BRDF measurements including human skin) containing human being's skin", Rendering Techniques'99 (Proceedings of the Eurographics Workshop on Rendering). It is patternized so that what is used by pp.131-144 and June 1999 may be resembled. By carrying out the tracking of these targets by the video sequence, the rigid body motion of the head is calculated, and after it, this is applied in order to carry out the rendering of the head model. Since only a video camera is required according to this setup, a motion of the head can be made demonstrating action of a wish.

[0081] In order to carry out the rendering of the instantiation system for carrying out the modeling of the reflection factor, and the face image with an approach real feeling, geometry is not only exact, but it is required that count of an echo of the light from the skin should be exact. In the example of a graphic display, in order to carry out the rendering of the face, the Monte Carlo beam-of-light marker of the physical base is used. As an example of technique, work besides Cook "distribution beam-of-light tracing (Distribution Ray Tracing)", Computer Graphics (SIGGRAPH'84 Proceedings), Work besides pp.165-174, July 1984, and Shirley "the Monte Carlo technique for direct-lighting count (Monte Carlo techniques for direct lighting calculations)", There are Transactions on Graphics, 15(1):1-36, and a thing indicated by 1996. According to this approach, BRDF (bi-directional reflectance distribution functions: bidirectional reflectance distribution function) of arbitration can be used, and approximation-ization can simulate the difficult appearance of the skin to accuracy with a simple shading model. Furthermore, although the extended light source is used, when carrying out the rendering of this like a portrait photograph, it is needed [ this ] in order to obtain the image which is not unsightly. He is trying for physical light migration to shift in the following two points on the need for count effectiveness. Namely, diffusion interreflection (diffuse interreflection) It is ignored, and the eye is illuminated so that it may pass without refraction along a pupil.

[0082] In the example of a graphic display, the reflection factor model of the skin is based on measurement of actual human being's face. As an example of technique, there are some which are indicated by work "the image base BRDF measurement (Image-based BRDF measurements including human skin) containing human being's skin" Rendering Techniques'99 (Proceedings of the Eurographics Workshop on Rendering) besides Marschner, pp.131-144, and June 1999. These measurement has described the average BRDF of the frame of some photographic subjects, and contains the adaptation parameter of the BRDF model indicated by work "the formation of nonlinear approximation of a reflection factor function (Non-linear approximation of reflectance functions)" besides Lafortune, Computer Graphics (SIGGRAPH'97 Proceedings), pp.117-126, and August 1997. Therefore, these measurement serves as a starting point which excelled when carrying out the rendering of the face with a real feeling. However, it is necessary to reinforce these measurement so that a part of spatial change observed by the actual face may be included. This is begun from the adaptation to the measurement BRDF of the skin in which the skin carried out the rendering, and a similar photographic subject, and is performed by dividing it into a diffuse reflection component and a specular reflection component. After it, in order to modulate each, the texture map is taken in.

[0083] According to the measurement taken from the actual face of a photographic subject, the texture map of a diffuse reflection component, i.e., an "albedo map", (albedo map) modulates a diffuse reflectance so that it may explain below. A specular reflection component is the Scala texture map. (scalartexture map) It is modulated, specular reflection is removed from the parts (eyebrows, eyelashes, etc.) which must not carry out a rendering at the rate of a cutaneous reflex, the specular reflection of the lower part of a face is reduced, and it is made for the description of the skin of a face to be resembled. The result is BRDF which changes spatially and this is work "the formation of nonlinear approximation of a reflection factor function (Non-linear approximation of reflectance functions)" besides Lafortune, Computer Graphics (SIGGRAPH'97 Proceedings), pp.117-126, and the conceptualization cosine lobe of August 1997. (generalized cosine lobes) It is described by each point by the sum.

[0084] The albedo map carried out the construction graphic display, and the albedo map on which it is required that the reflection factor resulting from diffuse reflection which changes spatially should be described was measured with the operation gestalt mentioned above using the digital photography sequence of the face photoed under the controlled lighting.

[0085] Drawing 11 shows the instantiation system used in order to capture digital photography or an image. In the system of a graphic display, a digital camera 1100 is equipped and two or more light sources are included. The example of whether to be and to attach is shown by 1102 and 1104. Since the polarizing filter which carried out the form of the vertical polarizers 1106, 1108, and 1110 was equipped and the light source and a camera lens are covered, specular reflection is inhibited and he is trying only for a diffuse reflection component to remain in an image. In this example, a photographic subject wears the hat 1112 with which the target which can be machine recognized was printed, and is made to carry out the tracking of the pause of the head. The camera 1110 remains standing it still, while the photographic subject was rotating. The only lighting has been obtained from the light sources 1102 and 1104 put on the measurement part near a camera. In order to reduce the indirect echo from the light which leaked, the black back drop (background curtain) is used.

[0086] Since it understands, the location of a camera and the light source is a field normal and irradiance. (irradiance) The coordinate to which it corresponds in an eye direction and texture space is calculable for every



pixel of each image using the standard beam-of-light tracing technique. Ideal lambert (Lambertian) If an echo is observed can calculate a lambert reflection factor from this information about the specific point in texture space under an assumption. If this count is repeated for every pixel contained in one photograph and that total is obtained, an image will be projected into texture space, and if it \*\* with the calculated irradiance resulting from the light source, the map of the diffuse reflection factor which crosses a field will be obtained. As shown in drawing 12, two photographs (A) and (B) are projected into texture space, and lighting is amended. Actually, this projection is reverse mapping repeated so that an outer loop may pass along all the pixels in a texture map. (reverse mapping) In order to take the average of the whole image area which is performed and is projected on a specific texture pixel, it is a probable super sampling. (stochastic supersampling) It is used.

[0087] The albedo map from a single photograph covers a part of field, and the result is grading. (grazing) It is best when an include angle is small. Therefore, in order to create the single albedo map of the whole face, the weighted average of all individual maps is calculated. Although the visual example is shown in drawing 13 (A) and (B), a weighting function is chosen so that high weight may be given to the pixel displayed and/or irradiated from the direction which is making the normal mostly to the field, and before either a display or an exposure carries out grading to the degree of pole, it descends to zero. The following function is used in the example of a graphic display.  $(\cos\theta_{\text{etai}} \cos\theta_{\text{etae}} - c)^p$  however  $c = 0.2$ ,  $p = 4$ .

[0088] A pixel is visible and irradiating suitably is confirmed, before calculating the albedo of a specific texture pixel. Even the camera point is traced with the point to the point on the light source on the pixel, and, as for two or more beams of light, a mark partial [ visibility and lighting ] in \*\* at zero or perfect is attached to the pixel. When estimating the size of the light source, it is wise to avoid a serious error. Only the albedo of the pixel which it is completely visible, is thoroughly irradiated by at least one light source, and is selectively irradiated by no light source is calculated. If it does in this way, the pixel in the pixel and all the shadows, or penumbra field hidden selectively will not be used.

[0089] Some calibration is required in order to give semantics to these measurement. The imprint curve of a camera is proofread using the approach indicated by work "recovery (Recovering high dynamic range radiance maps from photographs) of the high dynamic range radiance map from a photograph" besides Debevec, Computer Graphics (SIGGRAPH'97 Proceedings), pp.369-378, and August 1997. The flat field response of light / camera system is proofread using the photograph of the big White card. The focal length of a lens and a strain are proofread using Zhang work "the flexible new hand method for a camera calibration (A flexible new technique for camera calibration)", TechnicalReport MSR-TR -98-71, Microsoft Research, and the technique indicated by 1998. The whole scale factor is set using the reference sample of a known reflection factor. When change between the images of light source reinforcement is set as the object of consideration, control is performed by including a reference sample in all images.

[0090] The texture map obtained from the above-mentioned process is carrying out good work, when incorporating the detailed color change covering the whole face automatically. Furthermore, the strap used in order to hold a calibration hat in an orientation is visible. An image edit tool is used for these problems, and they are solved by embedding a texture or a homogeneity color in blank area soon.

[0091] Drawing 14 (A), (B), and drawing 15 (A) and (B) show the albedo map before edit, and the albedo map after edit for the purpose of a comparison. It is clear that be [ the part where an appropriate result is not obtained from an albedo map / where ] it is the part (for example, under a jaw) where a field is not fully observed, or is the part (for example, lug) which scans correctly and cannot be recorded with an image since the field is making the complicated configuration remainder. Neither of these parts need the texture from an albedo map, in order to acquire an appearance with a real feeling. It is because the 1st part is not visually conspicuous, and the 2nd part is because geometry is a detail in visual. Therefore, even if it edits these parts, most appearances of a final rendering are not influenced.

[0092] Drawing 16 - drawing 18 are drawings showing a different modality of some of face models which used the rest frame from the accompanying video. The face from some include angles is shown in the 1st ((A), (B), (C)) line. [ of drawing 16 ] There, it is shown how an albedo map and measured BRDF incorporate the appearance and color change of the skin covering the whole face seen from the include angle of arbitration. The 2nd ((A), (B), (C)) line shows the effectiveness of the lighting of a rim and side including the strong specular reflection of whenever [ grazing angle ]. [ of drawing 17 ] In addition, all of three images of this line are the same reinforcement, and the light source is put on the same distance from the face. Direction change of a reflection factor is contributing to a light effect with familiarity showing up in a rendering. By the 3rd ((A), (B), (C)) line, even if deformation of expression is added to a face, it is shown that a face seems to be natural in a motion of the usual expression. [ of drawing 18 ]

[0093] Drawing 19 is a flow chart which has indicated the step of the approach for creating an albedo map according to the operation gestalt currently explained here. This approach can be realized in suitable hardware, software, firmware, or such combination. With the operation gestalt currently explained here, this approach is realized by software as what is used by the same system as the system explained with reference to drawing 11.

[0094] One or the two polarization light sources or more which can be used at step 1700 since a photographic subject is irradiated are prepared. The example of the light source is as having mentioned above. With the operation gestalt currently explained here, the light source is chosen so that the specular reflection component of the face reflection factor of a photographic subject may be inhibited or removed. At step 1702, the face of a photographic subject is irradiated according to the light source. At step 1704, a photographic subject is rotated and a series of

digital photography or images are captured by the meantime. At step 1706, the coordinate in a field normal, irradiance, an eye direction, and texture space is calculated for every pixel in a texture map. This count is performed using a well-known algorithm. At step 1708, the lambert reflection factor to the specific pixel in the texture space of an image is calculated. Thereby, the albedo of a pixel is obtained. At step 1712, it is judged whether the additional pixel remains in the albedo map. If it remains, the following pixel will be obtained at step 1712 and it will return to step 1708. If the additional pixel does not remain in an albedo map, it is confirmed at step 1714 whether the additional digital image remains. If the additional digital image remains, the following digital image will be obtained at step 1716, and it will return to step 1706. If the additional digital image does not remain, at step 1718, the weighted average of the individual albedo map of each image is calculated, and the single albedo map of the whole face is created. The example of how this weighted average processing is performed is indicated by Marschner work "the reverse rendering (Inverse Rendering for Computer Graphics) of computer graphics", PhD thesis, Cornell University, and August 1998.

[0095] Drawing 20 is a flow chart which has indicated the step of the approach of calculating the albedo of a single pixel. This approach can be realized in suitable hardware, software, firmware, or such combination. This approach is realized by software with the operation gestalt currently explained here. At step 1800, when a certain pixel is given, it is judged whether the pixel is completely visible. The albedo of a pixel will not be calculated if a pixel is not completely visible (step 1804). If the pixel is completely visible, it will be judged at step 1802 whether the pixel is thoroughly irradiated by at least one light source. The albedo of a pixel is not calculated if the pixel is not thoroughly irradiated by at least one light source (step 1804). If the pixel is thoroughly irradiated by at least one light source, it will be judged at step 1806 whether the pixel is selectively irradiated by one of the light sources. If that is right, an albedo will not be calculated about the pixel. If the pixel is not selectively irradiated by one of the light sources, the albedo and weight of the pixel are calculated at step 1808. This weight packs an individual map later, and when taking that average, it is used. Therefore, as mentioned above, it is completely visible that an albedo is calculated and it is only about the pixel which is irradiated thoroughly and is not selectively irradiated according to one of the light sources by at least one light source. If it does in this way, the pixel in the pixel, all the shadows, or penumbra field hidden selectively will not be used.

[0096] As mentioned above, based on explanation of this description, if the application program for realizing the approach for the approach of animating the description of the explained face and expression conversion is this contractor, if it is this contractor, it can be created.

[0097] What is necessary is just to create an application program with the instruction which can perform CPU and in which read activation is possible. Moreover, one or two or more are [ that what is necessary is just to memorize the record medium which memorizes the above-mentioned application program to memory, a hard disk, and a well-known portable record medium ] sufficient as these record media. For example, the data used with application, for example, the data which defined a code book and training expression, are memorized to CDROM, and the application program which uses these data may store an application program and data in a record medium which is different as it memorizes to the hard disk. moreover, the thing which may be made to record on the same record medium — it is natural.

[0098] The computer which furthermore executes an application program may carry out distributed processing of the processing which one set is sufficient as and was specified with the application program by two or more computers.

[0099]

[Effect of the Invention] After meeting the high criteria that carry out the modeling of the face to a face with a real feeling, and it carries out a rendering to it according to the operation gestalt which has carry out conclusion \*\*\*\*, the system and approach of solving the difficult problem of enabling it to display in the form where the image which has familiarity like human being face is reliable are offer. The view of this approach is to use measurement as much as possible so that a face model may resemble an actual face actually. The geometry of a face is expressed in the fragmentation side with the connectability unified in common with a different face, and response nature mapped with the variation rate. It is obtained from BRDF measurement of the skin of human being by the former, and the measurement is used together with new measurement, and the reflection factor is combined so that two or more views (view) may become a single lighting amendment texture map showing a diffuse reflectance. A motion is obtained by the technique which was mentioned above and which is moved and incorporated, and the motion suitable for a shading side is made to be obtained by applying the motion to a face model using the improved deformation approach. The real feeling of a rendering is the renderer of the physical base about the geometry of an actual face, a motion, and a reflection factor. (physically-based renderer) It is improving substantially by using it.

[0100] As mentioned above, although it caught from the structural feature and/or the approach step and this invention has been explained, this invention clarified by the claim is not necessarily limited to the concrete description or approach step mentioned above so that I may be understood from the above explanation. Rather, the concrete description and an approach step are indicated as a desirable operation gestalt which realizes invention of a publication to a claim.

[Translation done.]

## \* NOTICES \*

JPO and NCIP I are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.\*\*\* shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

---

DESCRIPTION OF DRAWINGS

---

## [Brief Description of the Drawings]

[Drawing 1] It is the block diagram showing the general purpose computer suitable for using it when realizing the operation gestalt of this invention.

[Drawing 2] It is the block diagram showing the system which can be used in order to incorporate simultaneously the structure information and reflection factor information of a face on a photographic subject.

[Drawing 3] It is the flow chart which has indicated the instantiation approach for incorporating structure information and reflection factor information according to the operation gestalt of this invention.

[Drawing 4] It is the explanatory view showing the instantiation code book and transform function by the operation gestalt of this invention.

[Drawing 5] It is the flow chart which shows the expression translation process by the operation gestalt of this invention.

[Drawing 6] It is the block diagram showing the instantiation system which can adopt the principle of the operation gestalt of this invention of a certain kind.

[Drawing 7] (A) - (C) is drawing showing the collection of the instantiation collar plate in which the instantiation expression conversion by the operation gestalt of this invention is shown.

[Drawing 8] It is drawing showing the process which maps the same fragmentation control mesh in the fragmentation side by which the variation rate was carried out whenever faces differed.

[Drawing 9] (A) And (B) is drawing showing the instantiation constraint used in order to match the description in a field adaptation process.

[Drawing 10] It is the flow chart which has indicated the step of the field adaptation approach by the operation gestalt of this invention.

[Drawing 11] It is the block diagram showing an instantiation system employable in order to make the albedo map of a face according to the operation gestalt of this invention.

[Drawing 12] (A) And (B) is drawing showing the instantiation albedo map of two photographs with which it was projected into texture space and lighting was amended.

[Drawing 13] (A) And (B) is drawing showing the instantiation weighting function corresponding to the photograph shown in drawing 12.

[Drawing 14] (A) And (B) is drawing showing two perfect albedo maps to two different data sets.

[Drawing 15] (A) And (B) is drawing showing the albedo map of drawing 14 after editing.

[Drawing 16] (A) - (C) is drawing showing the collection of a face model by which the rendering was carried out to different sense under different lighting conditions.

[Drawing 17] (A) - (C) is drawing showing the collection of a face model by which the rendering was carried out to different sense under different lighting conditions.

[Drawing 18] (A) - (C) is drawing showing the collection of a face model by which the rendering was carried out to different sense under different lighting conditions.

[Drawing 19] It is the flow chart which has indicated the step of the approach for creating an albedo map according to the operation gestalt of this invention.

[Drawing 20] It is the flow chart which has indicated the step of the approach for calculating the albedo of a single pixel according to the operation gestalt of this invention.

## [Description of Notations]

200 Lighting System

202 Camera

202a The lens of a camera

204 206 Light source

400 Code Book

402 Training Expression Set

406 Conversion Processor

408 Transform Function

410 Synthetic Expression Set

602 Transmitter (Transmitting Side)

604 Receiver (Receiving Side)

606 Reconstruction Module

---

[Translation done.]

## \* NOTICES \*

JPO and NCIP are not responsible for any damages caused by the use of this translation.

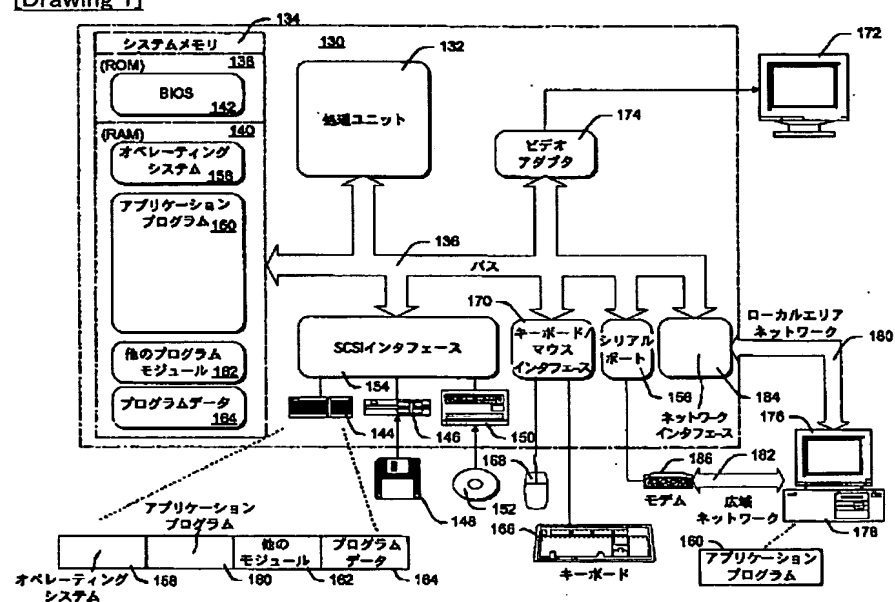
1.This document has been translated by computer. So the translation may not reflect the original precisely.

2.\*\*\* shows the word which can not be translated.

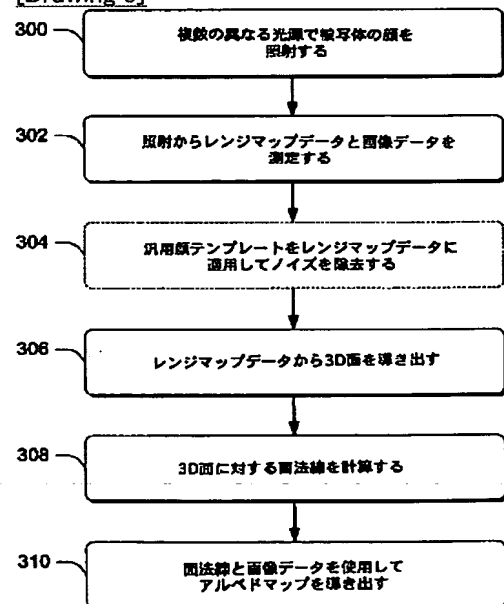
3.In the drawings, any words are not translated.

## DRAWINGS

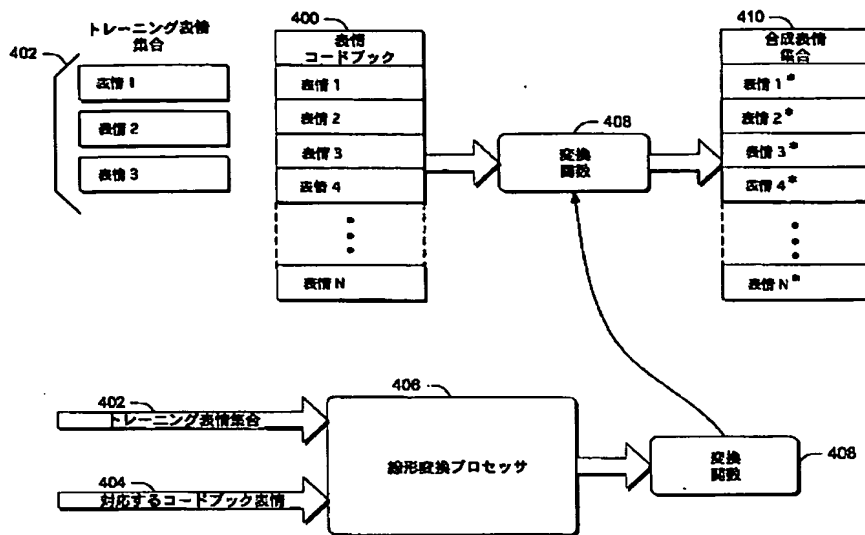
[Drawing 1]



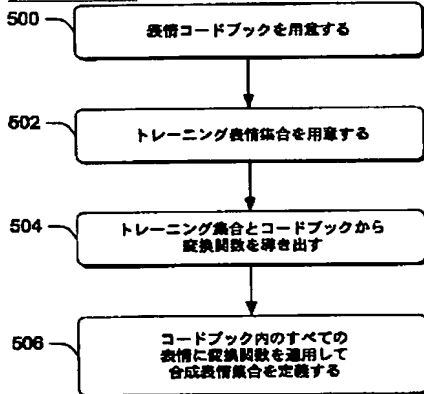
[Drawing 3]



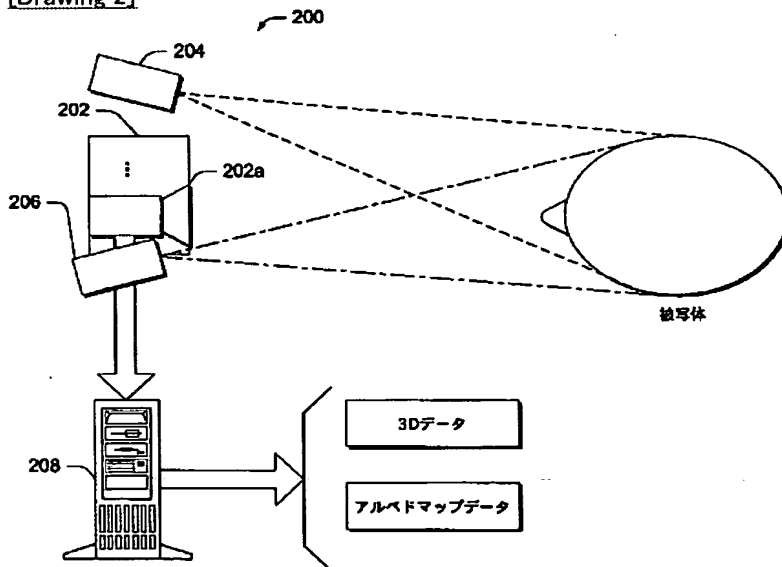
[Drawing 4]



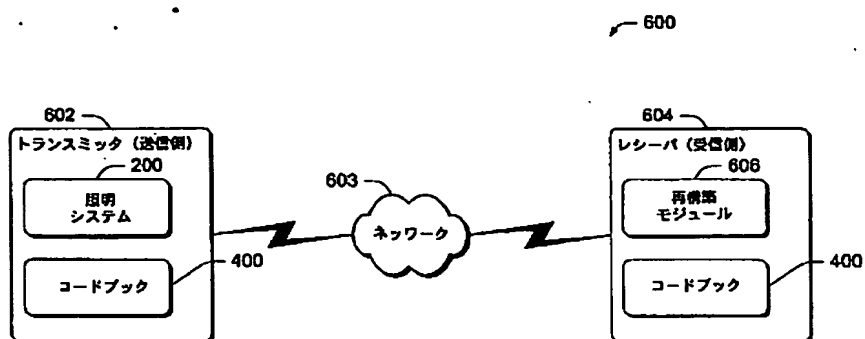
[Drawing 5]



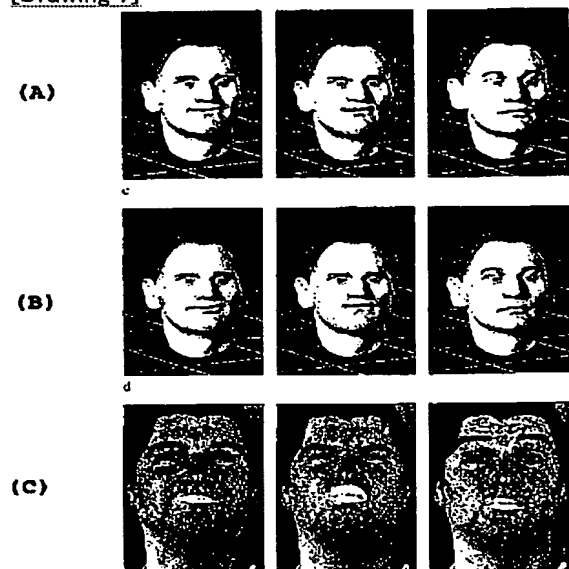
[Drawing 2]



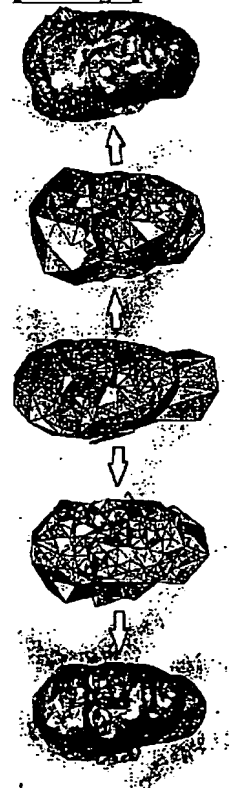
[Drawing 6]



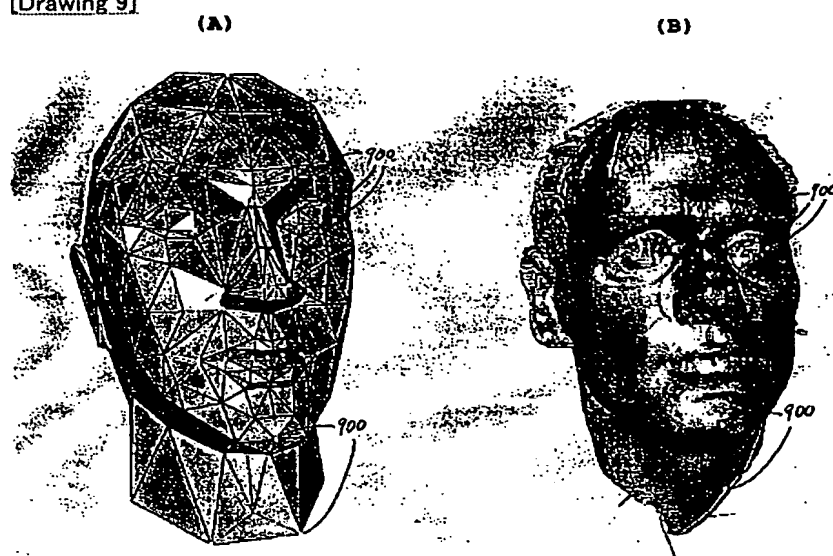
[Drawing 7]



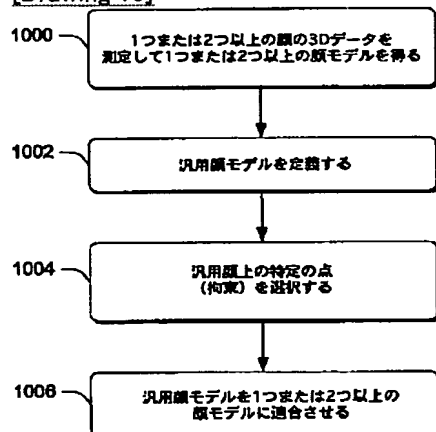
[Drawing 8]



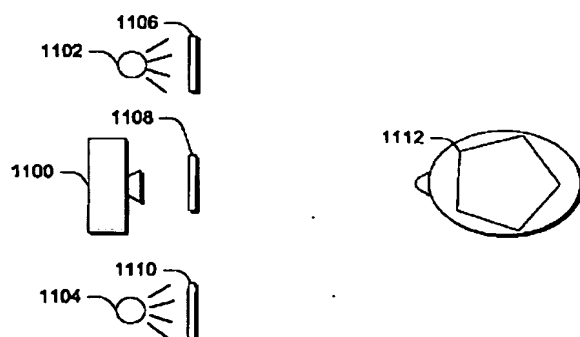
[Drawing 9]



[Drawing 10]



[Drawing 11]



[Drawing 12]



(A)

(B)

[Drawing 13]

(A)

(B)

[Drawing 16]

(A)

(B)

(C)



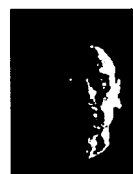
1行目

[Drawing 17]

(A)

(B)

(C)



2行目

[Drawing 18]

(A)

(B)

(C)



3行目

[Drawing 14]

(A)

(B)



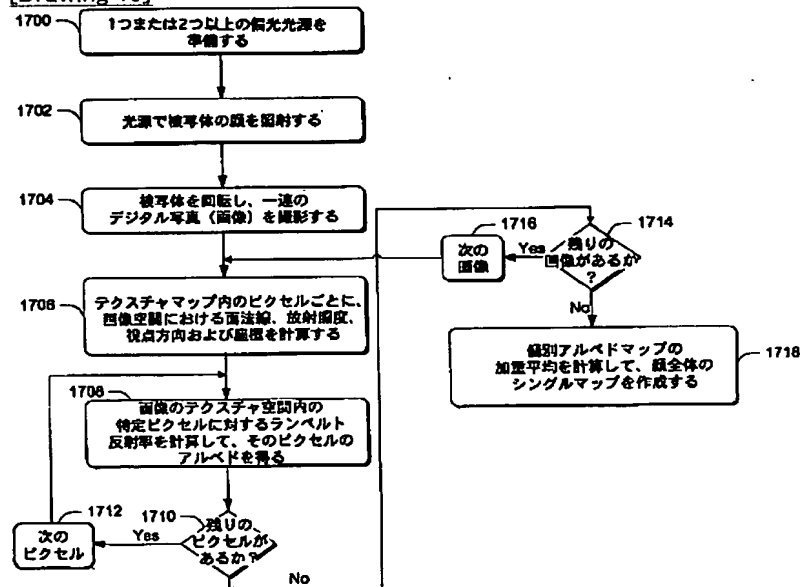
[Drawing 15]

(A)

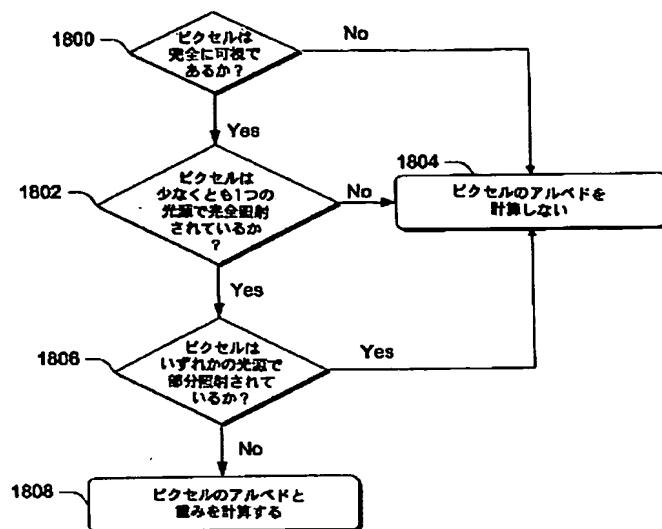
(B)



[Drawing 19]



[Drawing 20]



[Translation done.]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2002-123837

(P2002-123837A)

(43)公開日 平成14年4月26日(2002.4.26)

(51)Int.Cl. <sup>7</sup>		識別記号	F I		テーマコード <sup>8</sup> (参考)	
G 0 6 T	15/70		G 0 6 T	15/70	B	5 B 0 5 0
	1/00	3 4 0		1/00	3 4 0 A	5 B 0 5 7
	7/20	3 0 0		7/20	3 0 0 B	5 L 0 9 6

審査請求 未請求 請求項の数51 O L (全 23 頁)

(21)出願番号	特願2001-261558(P2001-261558)	(71)出願人	391055933 マイクロソフト コーポレーション MICROSOFT CORPORATION アメリカ合衆国 ワシントン州 98052-6399 レッドモンド ワン マイクロソフト ウェイ (番地なし)
(22)出願日	平成13年8月30日(2001.8.30)	(72)発明者	スティーブン マーシュナー アメリカ合衆国 98033 ワシントン州 カークランド 21 プレイス 11
(31)優先権主張番号	09/651880	(74)代理人	100077481 弁理士 谷 義一 (外2名)
(32)優先日	平成12年8月30日(2000.8.30)		
(33)優先権主張国	米国(US)		

最終頁に続

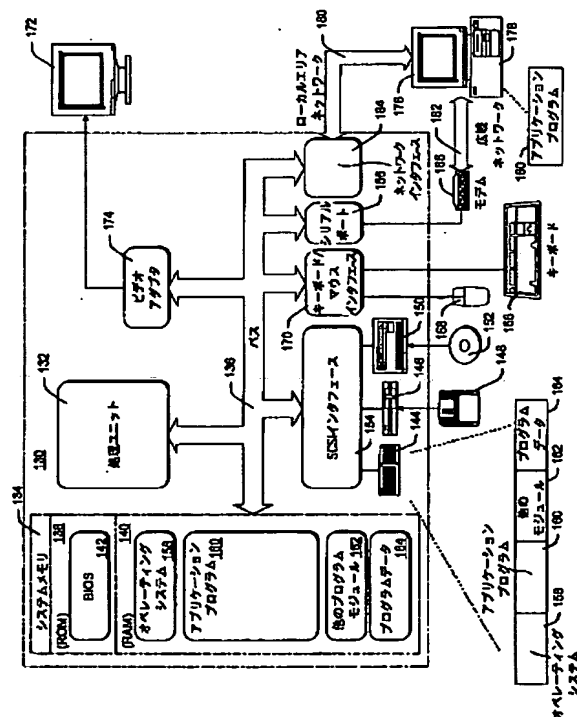
**最終頁に続く**

(54) 【発明の名称】 顔の特徴をアニメーション化する方法およびシステムならびに表情変換のための方法およびシステム

(57) 【要約】

【課題】 顔の特徴を正確にアニメーション化し、顔の表情を変換するための方法およびシステムを提供する。

【解決手段】 コードブックは、第1人物の顔表情集合を定義しているデータを収めている。第2人物からの顔トレーニング表情集合およびコードブックからの対応する表情は変換関数を導き出すために使用され、この変換関数はコードブックの表情のすべてに適用されるようになってい。このようにして、第1人物からの表情は、第2人物の表情にリアル感をもって変換すること、およびその逆に変換することを可能にしている。多数の異なる顔を適合させる操作の基礎として使用される、単一共通汎用顔モデルが用意されている。この単一共通汎用顔モデルとユーザが定義したある種の制約を使用することにより、異なる顔間の対応付けを可能にするメカニズムが得らる。これらの変換処理は処理ユニット132で実行される。



## 【特許請求の範囲】

【請求項 1】 第 1 人物の第 1 顔表情集合を定義しているデータを収めているコードブックを定義し、第 1 人物とは異なる第 2 人物のトレーニング表情集合を与える第 2 顔表情集合を定義しているデータを用意し、前記トレーニング表情集合と、前記第 1 表情集合からの対応する表情とから変換関数を導き出し、前記変換関数を該第 1 表情集合に適用して合成表情集合を得ることを特徴とする顔表情変換方法。

【請求項 2】 請求項 1 に記載の方法において、前記トレーニング表情集合は、コードブックよりも少ない表情を含んでいることを特徴とする方法。

【請求項 3】 請求項 1 に記載の方法において、前記変換関数は、第 1 人物および第 2 人物の顔のサイズと形状の差異を補償することを特徴とする方法。

【請求項 4】 請求項 1 に記載の方法において、前記変換関数を導き出すステップでは、一方の表情集合から他方の表情集合への線形的変換を計算することを特徴とする方法。

【請求項 5】 請求項 1 に記載の方法において、前記変換関数を導き出すステップでは、  
 各々の表情を、 $m$  個の標準サンプル位置での  $x$ ,  $y$ ,  $z$  変位を含んでいる  $3m$  ベクトルとして表現し、  
 変換される顔の  $n$  個の表情ベクトル集合が  $g_{a1} \dots g_{an}$  として与えられ、ターゲット顔の、対応するベクトル集合が  $g_{b1} \dots g_{bn}$  として与えられているとき、次の形式の  $3m$  線形最小二乗システム、

$$a_j \cdot g_a = g_b[j], i = 1 \dots n$$

を解くことによって、 $g_a$  の各々の座標ごとに 1 つの、線形予測子集合  $a_j$  を計算することを特徴とする方法。

【請求項 6】 請求項 5 に記載の方法において、前記計算するステップでは、各々の  $g_{aj}$  についてポイントの部分集合だけを使用することを特徴とする方法。

【請求項 7】 請求項 6 に記載の方法において、前記使用するステップでは、考慮の対象となっている標準サンプルポイントとエッジを共有しているポイントだけを使用することを特徴とする方法。

【請求項 8】 請求項 5 に記載の方法において、さらに、 $a_j$  の解を求めるために擬似逆関数を計算するとき、特異値の拡散を制御することを特徴とする方法。

【請求項 9】 請求項 8 に記載の方法において、前記拡散を制御するステップは、 $\alpha \sigma_1$  未満のすべての特異値をゼロとなし、ここで、 $\sigma_1$  は、マトリックスの最大特異値であることを特徴とする方法。

【請求項 10】 請求項 1 に記載の方法において、前記第 2 顔表情集合を定義しているデータを用意するステップでは、  
 第 2 人物の顔を照明で照射し、

顔の構造を記述している構造データと、照射からの顔の反射率特性を記述している反射率データとを同時に取り込む、

ことを特徴とする方法。

【請求項 11】 請求項 10 に記載の方法において、前記照射するステップでは、  
 複数の光源を使用し、それらの光源の 1 つは第 2 人物の顔上にパターンを投影し、そのパターンから構造データを確可能となし、

10 前記光源の少なくとも 1 つは赤外線光源であり、

該光源の少なくとも 1 つは偏光されており、

前記取り込むステップでは、鏡面反射光を抑止して拡散成分反射データを取り込むポラライザを備えたカメラを使用することを特徴とする方法。

【請求項 12】 請求項 1 に記載の方法において、第 2 顔表情集合を定義しているデータを用意するステップでは、

顔の鏡面反射抑止反射率特性を確認できるように選択された第 1 偏光光源で第 2 人物の顔を照射し、

20 第 1 偏光光源で顔を照射している間に、その照射と同時に、顔上にパターンを投影する第 2 構造化光源で第 2 人物の顔を照射し、

同時照射から鏡面反射抑止反射データと構造データの両方を取り込むことを特徴とする方法。

【請求項 13】 請求項 12 に記載の方法において、前記第 1 偏光光源および前記第 2 構造化光源は異なる周波数の光を出力することを特徴とする方法。

【請求項 14】 請求項 12 に記載の方法において、前記第 1 偏光光源および前記第 2 構造化光源は赤外線光を出力することを特徴とする方法。  
 30

【請求項 15】 請求項 12 に記載の方法において、さらに、取り込まれたデータを処理して、(a) 顔の次元様相を記述しているデータと、(b) 顔の拡散反射特性を記述しているデータの両方を得るようにしたことを特徴とする方法。

【請求項 16】 請求項 1 に記載の方法において、前記第 2 顔表情集合を定義しているデータを用意するステップでは、

複数の異なる光源で第 2 人物の顔を照射し、

40 前記照射からレンジマップデータを測定し、

該照射から画像データを測定し、

前記レンジマップデータから 3 次元面を導き出し、

前記 3 次元面に対する面法線を計算し、

前記面法線と前記画像データを処理してアルベドマップを導き出すことを特徴とする方法。

【請求項 17】 請求項 16 に記載の方法において、複数の光源の少なくとも 1 つは偏光されていることを特徴とする方法。

【請求項 18】 請求項 16 に記載の方法において、複数の光源のすべては偏光されていることを特徴とする方  
 50

法。

【請求項 19】 コンピュータによって実行されるとき、ある人物からのトレーニング表情集合と、別の人物のコードブックからの、対応する表情とを操作し、前記トレーニング表情集合と、これらに対応する表情とから線形的変換を計算するステップ、コードブックからの複数の表情に変換関数を適用して合成表情集合を得るステップと、をコンピュータに実行させるためのコンピュータ読取可能命令を格納したことを特徴とする 1 つまたは 2 つ以上のコンピュータ読取可能媒体。

【請求項 20】 請求項 19 に記載のコンピュータ読取可能媒体において、前記コンピュータ読取可能命令は、コンピュータに合成表情集合を使用させて、一方の人物からの表情を他方の人物の表情に変換させることを特徴とするコンピュータ読取可能媒体。

【請求項 21】 請求項 20 に記載のコンピュータ読取可能媒体において、前記コンピュータ読取可能命令は、一方の人物からの表情であって、コードブックの表情を構成している表情とは異なっている表情を、コンピュータに変換させることを特徴とするコンピュータ読取可能媒体。

【請求項 22】 請求項 20 に記載のコンピュータ読取可能媒体において、前記コンピュータ読取可能命令は、合成表情の少なくとも 1 つのインデックスを、表情を再構築できるレシーバに送信することによって、コンピュータに表情を変換させることを特徴とするコンピュータ読取可能媒体。

【請求項 23】 請求項 20 に記載のコンピュータ読取可能媒体において、前記コンピュータ読取可能命令は、コンピュータに顔の表情を変換させることを特徴とするコンピュータ読取可能媒体。

【請求項 24】 コンピュータ読取可能媒体上に具現化されたコードブックであって、第 1 人物の第 1 顔表情集合を定義しているデータを収めているコードブックと、コンピュータ読取可能媒体上に具現化されたデータであって、該データは第 2 顔表情集合を定義しており、該第 2 顔表情集合は、前記第 1 人物とは異なる第 2 人物のトレーニング表情集合となっているデータと、前記トレーニング表情集合と、前記第 1 表情集合からの対応する表情とから変換関数を導き出す変換プロセッサとを備えたことを特徴とする顔表情変換システム。

【請求項 25】 請求項 24 に記載の顔表情変換システムにおいて、前記変換プロセッサは線形変換プロセッサを有することを特徴とする顔表情変換システム。

【請求項 26】 請求項 24 に記載の顔表情変換システムにおいて、さらに、コンピュータ読取可能媒体上に具現化された合成表情集合を含み、該合成表情集合は、変換関数をコードブック表情に適用することによって導き

出されることを特徴とする顔表情変換システム。

【請求項 27】 請求項 24 に記載の顔表情変換システムにおいて、前記変換関数は、第 1 人物および第 2 人物の顔のサイズと形状の差異を補償することを特徴とする顔表情変換システム。

【請求項 28】 請求項 24 に記載の顔表情変換システムにおいて、前記変換プロセッサは、各々の表情を、 $m$  個の標準サンプル位置での  $x$ ,  $y$ ,  $z$  変位を含んでいる  $3m$  ベクトルとして表現し、

10 変換される顔の  $n$  個の表情ベクトル集合が  $g_{a1} \dots g_{an}$  として、ターゲット顔の、対応するベクトル集合が  $g_{b1} \dots g_{bn}$  として与えられているとき、次の形式の  $3m$  線形最小二乗システム、

【数 2】

$$a_j \cdot g_a = g_b[j], i = 1 \dots n$$

を解くことによって、 $g_a$  の各々の座標ごとに 1 つの、線形予測子集合  $a_j$  を計算することによって変換関数を導き出すことを特徴とする顔表情変換システム。

20 【請求項 29】 被写体の顔を照射するために同時に複数の異なる光源を提供する顔照明システムと、前記顔照明システムによって照射されたとき、被写体の顔から構造データと反射率データの両方を取り込むデータ取り込みシステムと、第 1 合成表情コードブックであって、該第 1 合成表情コードブックは、被写体から得られたトレーニング表情集合を受信し、前記トレーニング表情集合と、対応する未合成コードブック表情とを使用して変換関数を計算し、前記変換関数を第 1 合成情報コードブック内の表情のすべてに適用することによって合成される第 1 合成情報コードブックとを有するトランスミッタと、

30 該トランスミッタと通信可能にリンクされたレシーバであって、顔画像を再構築するための再構築モジュールと、前記第 1 合成情報コードブックと同じ合成表情を収めている第 2 合成情報コードブックとを有するレシーバとを備え、前記トランスミッタは、被写体の追加表情を取り込み、対応するまたはほぼ合致する表情を探すために第 1 コードブックをサーチし、対応するまたはほぼ合致するコードブック表情のインデックスを、再構築モジュールによる顔画像再構築のために前記レシーバに送信することを特徴とする顔表情変換システム。

40 【請求項 30】 請求項 29 に記載の顔表情変換システムにおいて、前記照明システムは、少なくとも 1 つの偏光光源を含むことを特徴とする顔表情変換システム。

【請求項 31】 請求項 29 に記載の顔表情変換システムにおいて、前記照明システムは、複数の偏光光源を含むことを特徴とする顔表情変換システム。

50 【請求項 32】 請求項 29 に記載の顔表情変換システムにおいて、前記照明システムは、被写体の顔の上にパターンを投影するように構成されたパターン化光源を含

## 5

むことを特徴とする顔表情変換システム。

【請求項 3 3】 請求項 2 9 に記載の顔表情変換システムにおいて、前記照明システムは、被写体の顔の上にパターンを投影する赤外線光源を含むことを特徴とする顔表情変換システム。

【請求項 3 4】 請求項 2 9 に記載の顔表情変換システムにおいて、前記複数の異なる光源は、すべてが赤外線光源であることを特徴とする顔表情変換システム。

【請求項 3 5】 顔の特徴をアニメーション化する方法であって、  
複数の異なる顔のジオメトリに近似している細分面を定義し、

複数の顔の各々に同じ細分面を適合させることを特徴とする顔の特徴をアニメーション化する方法。

【請求項 3 6】 請求項 3 5 に記載の方法において、前記定義するステップでは、粗メッシュ構造をもつ細分面を定義することを特徴とする方法。

【請求項 3 7】 請求項 3 6 に記載の方法において、粗メッシュ構造は三角メッシュを有することを特徴とする方法。

【請求項 3 8】 請求項 3 5 に記載の方法において、前記適合させるステップでは、細分面の頂点位置にわたって連続最適化オペレーションを実行することを特徴とする方法。

【請求項 3 9】 請求項 3 5 に記載の方法において、前記適合させるステップでは、細分面を定義しているメッシュの接続性を変更することなく、細分面を顔に適合させることを特徴とする方法。

【請求項 4 0】 請求項 3 5 に記載の方法において、前記適合させるステップでは、細分面を定義しているメッシュに関連する平滑化汎関数を最小限にすることを特徴とする方法。

【請求項 4 1】 請求項 3 5 に記載の方法において、前記適合させるステップでは、細分面を定義しているメッシュに関連する 1 つまたは 2 つ以上の制約を選択し、前記制約を、面上の対応するポイントに直接に適合させることを特徴とする方法。

【請求項 4 2】 請求項 4 1 に記載の方法において、前記制約は、目、鼻および口の 1 つに関連付けられていることを特徴とする方法。

【請求項 4 3】 請求項 3 5 に記載の方法において、前記適合させるステップでは、距離、滑らかさ、および制約を表す項を含む汎関数を最小限にすることを特徴とする方法。

【請求項 4 4】 請求項 3 5 に記載の方法において、前記適合させるステップでは、一連の線形最小二乗問題の解を求めることを特徴とする方法。

【請求項 4 5】 1 つまたは 2 つ以上のコンピュータによって実行されたとき、当該 1 つまたは 2 つ以上のコンピュータに請求項 3 5 に記載の方法を実現させるコンピ

(4)

特開 2002-123837

## 6

ュータ読取可能命令を格納したことを特徴とする 1 つまたは 2 つ以上のコンピュータ読取可能媒体。

【請求項 4 6】 顔の特徴をアニメーション化する方法であって、

複数の異なる顔のジオメトリに近似する細分面を定義し、

複数の前記顔の各々に同じ細分面を適合させて、顔と顔とを対応付け、

顔と顔との対応付けを使用して一方の顔の表情を他方の顔の表情に変換することを特徴とする顔の特徴をアニメーション化する方法。

【請求項 4 7】 顔の特徴をアニメーション化する方法であって、

複数の異なる顔について 3 次元データを測定して、対応する顔モデルを取得し、

各々の対応する顔モデルにマッピングするために使用される汎用顔モデルを 1 つだけ定義し、

対応する各々の顔モデル上の対応するポイントに直接にマッピングされる、前記汎用顔モデル上の複数のポイントを選択し、

対応する顔モデルの各々に該汎用顔モデルを適合させ、当該適合させるステップでは、選択されたポイントの各々を、対応する各々の顔モデル上の、対応するポイントに直接にマッピングすることを特徴とする顔の特徴をアニメーション化する方法。

【請求項 4 8】 請求項 4 7 に記載の方法において、前記定義するステップでは、ベースメッシュ構造から細分面を定義し、前記細分面は複数の頂点を含み、顔モデルのジオメトリを近似化しており、前記適合させるステップでは、細分面の頂点の位置だけを操作することを特徴とする方法。

【請求項 4 9】 請求項 4 7 に記載の方法において、前記適合させるステップでは、細分面を定義しているベースメッシュを操作することを特徴とする方法。

【請求項 5 0】 請求項 4 7 に記載の方法において、前記適合させるステップでは、細分面を定義しているベースメッシュを、ベースメッシュの接続性を変更することなく操作することを特徴とする方法。

【請求項 5 1】 請求項 4 7 に記載の方法において、前記測定するステップでは、レーザレンジスキャンを使用して 3 次元データを測定することを特徴とする方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、顔をリアルにアニメーション化するためのモデリングとレンダリングの方法およびシステムに関する。具体的には、本発明は顔画像処理のための方法およびシステムに関する。

【0002】

【従来の技術】コンピュータグラフィックスの分野では、さまざまなオブジェクトをレンダリングして、その

## 7

オブジェクトがユーザのためにコンピュータディスプレイから表示される。例えば、コンピュータゲームでは、コンピュータオブジェクトを生成し、レンダリングしてコンピュータモニタやテレビジョンに表示するためのコンピュータグラフィックスアプリケーションを含むことが代表的である。リアル感のある画像のモデリングとレンダリングは、コンピュータグラフィックス分野に従事するものにとっては、絶え間のない努力目標となっている。コンピュータグラフィックス分野で特に努力目標となっている分野は、リアル感のある顔画像をレンダリングすることである。1つの例として、ある特定のコンピュータグラフィックスアプリケーションでは、会話に参加している個人の表示をレンダリングすることが行われている。よく見られることは、この個人の、最終的にレンダリングされた画像はコンピュータでレンダリングされた画像であり、現実の個人とはしばしば大幅に異なっている。

## 【0003】

【発明が解決しようとする課題】リアル感のある顔と顔表情をモデリングし、レンダリングすることは、特に2つの理由で困難な作業となっている。1つは、人間の皮膚が、現在利用可能である、種々のシェーディングモデルでモデリングするのに不適切な反射率特性をもっているからである。例えば、周知のPhongモデルは、人間の皮膚をモデリングするには不十分である。もう1つは、顔表情をレンダリングするとき、「リアルな」顔の動きとして感知されるものと若干でもずれがあると、そのずれは、無頓着な観察者でさえも、正しくないものとして感知されるからである。現存の、顔の動き取り込み（キャプチャ）システムを使用すると、かなり説得力のある顔アニメーションを作成することができるが、取り込まれた動きは、別の顔に適用されたときは、その説得力がはるかに低下し、非常に奇異になることがよく起こっている。例えば、ある人物からその顔の動きのサンプリングをとれば、個々の顔の動きをアニメーション化することは、その動きの元となった顔が同じ顔であることを考慮すれば、困難ではない。その結果、ある表情と別の表情の間で同じか、あるいは非常に類似している動きの特徴が得られることになる。しかし、この人物の顔の動きを別の人物の顔に変換することは、この2つの顔の間には固有の差異（例えば、顔のサイズと形状）があるため、説得力に欠けることがよく起こっている。

【0004】従って、本発明は、人間の皮膚のテクスチャと反射率をモデリングするための改良システムと方法を提供することを課題にしてなされたものである。また、本発明は、ある人物の顔の動きを別人物の顔の動きに変換することによって、顔の動きの取り込み（キャプチャ）データを再使用するためのシステムと方法を提供することを課題にしてなされたものである。

## 【0005】

## 8

【課題を解決するための手段】以下に図示し、説明している実施形態では、3次元（3-D）顔つきを記述しているデータを取り込んで、顔の動きを、ある個人から別の個人にリアルに変換し、皮膚の反射率をモデリングするための、新規な手法が提案されている。

【0006】ここで説明している実施形態では、人間の被写体が用意され、被写体の顔を照射するために複数の異なる光源が利用されている。複数の光源の中の1つは、被写体の顔の上にパターンを投影する構造化光源（structured light source）である。この構造化光源により1つまたは2つ以上のカメラが、被写体の3-Dの顔つきを記述しているデータを取り込むことが可能となる。他の光源がさらにもう1つ用意されているが、これは被写体の顔を照射するために使用されるものである。この他の1つの光源により被写体の顔の種々反射率特性（reflectance properties）を確認する。また、この他方の光源は偏光フィルタ（polarizing filter）と併用されて、顔の反射率の鏡面反射成分（specular component）が除去される。すなわち、拡散反射成分（diffuse component）だけがカメラによって取り込まれる。複数の異なる光源を使用すると、顔の構造特性と反射率特性の両方を、同時に確認することが可能になる。光源の選択を慎重に行うと、例えば、光源をナローバンドにし、整合ナローバンドフィルタ（matching narrowband filter）をカメラで使用すると、周囲照明源の影響を除去することが可能になる。

【0007】ここで説明している照明プロセスの結果として、2つの有用なアイテムが得られる。すなわち、(1) レンジマップ（またはデプスマップ（depth map））と、(2) 構造化光源のパターンがそこに投影されていない顔の画像である。3D面がレンジマップから導き出され、3D面に対する面法線（surface normal）が計算される。3D面を定義するためにレンジマップを処理するとき、その処理にオプションとしてのフィルタリングステップを含めておくと、汎用顔テンプレート（generic face template）がレンジマップと結合され、不要なノイズが拒絶されることになる。計算された面法線と顔の画像は、アルベドマップ（albedo map）を導き出すために使用される。アルベドマップは特殊タイプのテクスチャマップ（texture map）であり、そこでは、各サンプルは、面上の特定ポイント（点）における顔面の拡散反射率を記述している。従って、プロセスのこの時点で、顔の3D面（つまり、面法線）を記述している情報と、顔の反射率（つまり、アルベドマップ）を記述している情報は確認されていることになる。

【0008】ある実施形態では、照明プロセスで得られた情報またはデータは、ある人物の顔の表情を、別人物の顔の表情に変換するために使用される。この実施形態では、コードブック（code book）の考え方が取り入れられ、使用されている。



【0009】コードブックは、別人物（人物A）の、多数の汎用表情（generic expressions）を記述しているデータを収めている。1つの目標は、コードブックの表情を取り出し、その表情を使用して別人物（人物B）の表情を変換することである。これを行うために、本発明による方法では、人物Bを使用してトレーニング表情（training expressions）の集合を作成している。トレーニング表情は、コードブックに存在する表情の集合から構成されている。トレーニング表情および各表情に対応するコードブック表情を使用すると、変換関数（transformation function）が導き出される。この変換関数は、人物Bの表情に合致しているはずの合成表情の集合（a set of synthetic expressions）を導き出すために使用される。すなわち、変換関数が導き出されると、これはコードブック内の表情の各々に適用され、コードブック表情が人物Bの表情に合致するようにされる。従って、トレーニング集合の中に存在しない可能性のある、新しい表情が、例えば、人物Bから受信されたときは、合成コードブック表情をサーチすれば、人物Bの表情に最も合致する表情を見付けることができる。

【0010】別の実施形態では、顔の表情と動きを、ある顔から別の顔に変換するために使用できる共通顔構造（common face structure）が定義されている。ここで説明している実施形態では、共通顔構造は、粗メッシュ構造（coarse mesh structure）または「ベースメッシュ（base mesh）」からなり、これは、ある人物の表情を別の人物の表情に変換するときの基礎として使用される細分面（subdivision surface）を定義している。共通ベースメッシュはすべての顔に対して使用され、2または3以上の顔が対応付けられるようにしている。従って、これによって定義された構造を使用すると、顔の動きをある人物から別人物に適応することが可能になる。この実施形態によれば、細分面を被写体の顔モデルに適応するための手法が用いられている。本発明による手法では、細分面上の、いくつかのポイントは、顔モデル上の対応するポイントに直接的にマッピングされるように定義されている。このことは、起こり得る、あらゆる異なる顔モデルについても同じである。この制約（constraint）を加えると、ベースメッシュは、異なる顔モデルを同じように適合させる、という特性をもつことになる。さらに、本発明によるアルゴリズムでは、ベースメッシュと顔モデルとが確実に対応づけられるように最小限にされた平滑化汎関数（smoothing functional）が利用されている。

【0011】別の実施形態では、被写体の顔面の反射率を計量化する反射率処理手法が用意されている。反射率を測定するために、本発明による手法では、反射率は拡散反射成分（diffuse component）と鏡面反射成分（specular component）に分離され、拡散反射成分の処理が中心になっている。

【0012】拡散反射成分を測定するために、アルベドマップ（albedo map）が最初に定義される。このアルベドマップは、カメラと、複数の異なる光源によって照射される被写体とを、最初に用意することによって定義される。光源は偏光フィルタ（polarizing filter）によってフィルタリングされる。このフィルタリングは、偏光フィルタがカメラの前に置かれた偏光フィルタと併用されて、鏡面反射を抑止するか、あるいは鏡面反射が記録されるのを防止することによって行われる。画像シーケンスは、被写体の頭を取り巻くように撮影される。各々の個別画像は、その画像に対応する個別アルベドマップが得られるように処理される。特定の被写体に対するすべてのアルベドマップは、その被写体の顔全体に対して単一のアルベドマップが得られるように結合される。

#### 【0013】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を詳細に説明する。

#### 【0014】概要

リアル感のある顔と顔表情をレンダリングするには、皮膚の反射率と顔の動きを非常に正確なモデルで表す必要がある。以下では、ジオメトリ、動きおよび反射率に関する測定データを使用して、顔をモデリングし、アニメーション化し、レンダリングして特定人物の顔の外観と顔表情をリアルに再現するための方法および手法が説明されている。ジオメトリと双方向反射率を含んでいる完全モデルが構築されるので、どのような照明条件と表示条件においても、顔をレンダリングすることができる。以下に説明するモデリングシステムと方法では、異なる顔間が対応付けられた構造化顔モデル（structured face model）が構築され、この顔モデルは、さまざまな顔のアニメーション操作の基礎となっている。

【0015】本発明による実施形態は、以下に説明するように、顔モデリングプロセスを構成する各部分と係わりをもっている。ジオメトリの表現を構造化し、統一化することによって、そのジオメトリ表現を顔モデルの基礎とし、多くの以後の顔モデリングとレンダリング操作の基礎となるようにするために、本発明の実施形態では、従来の面適合手法（surface fitting technique）を拡張し、汎用の顔（generic face）が異なる個人の顔に合致させるようにしている。リアル感のある反射率モデルを構築するために、最初に、最近の皮膚反射率測定が使用されるが、これは公知で実用化されている。さらに、改良されたテクスチャ取り込みプロセス（texture capture process）を使用して、新規に測定された拡散テクスチャマップが付加されている。汎用メッシュ（generic mesh）をアニメーション化するために、改良された手法が使用され、高品質レンダリングに適した面形状が得られるようにしている。

#### 【0016】例示コンピュータシステム

予備知識として、図1に、ここで説明している実施形態

で利用できるデスクトップコンピュータ130の例の概要を示す。分散コンピューティング環境の状況下では、図示のようなコンピュータはさまざまなものが使用可能である。これらのコンピュータは、以下に記載されている説明に従ってグラフィックスをレンダリングし、画像（イメージ）を処理するために使用することができる。

【0017】コンピュータ130は、1つまたは2つ以上のプロセッサまたは処理ユニット132、システムメモリ134、およびバス136、すなわち、システムメモリ134を含む種々のシステムコンポーネントをプロセッサ132に結合しているバス136を含んでいる。バス136は、数種タイプのバス構造のいずれか1つまたは2つ以上で表されているが、この中には、さまざまなバスアーキテクチャを採用したメモリバスまたはメモリコントローラ、ペリフェラルバス、高速グラフィックスポート、およびプロセッサまたはローカルバスが含まれている。システムメモリ134としては、リードオンリメモリ（read only memory - ROM）138とランダムアクセスメモリ（random access memory - RAM）140がある。基本入出力システム（basic input/output system - BIOS）142は、スタートアップ時のときに、コンピュータ内のエレメント間の情報転送を支援する基本ルーチンで構成され、ROM 138に格納されている。

【0018】コンピュータ130は、さらに、ハードディスク（図示せず）との間で読み書きするハードディスクドライブ144、取り外し可能ディスク148との間で読み書きする磁気ディスクドライブ146、CD-ROMや他の光媒体のような、取り外し可能光ディスク152との間で読み書きする光ディスクドライブ150を装備している。ハードディスクドライブ144、磁気ディスクドライブ146、および光ディスクドライブ150は、SCSIインタフェース154または他の、ある種の該当ペリフェラルインタフェースを通してバス136に接続されている。これらのドライブおよびその関連コンピュータ読取可能媒体は不揮発性ストレージとして、コンピュータ読取可能命令、データ構造、プログラムモジュールおよび他のデータを、コンピュータ130のために格納している。ここで説明している例示環境では、ハードディスク、取り外し可能磁気ディスク148および取り外し可能光ディスク152が採用されているが、当業者ならば理解されるように、磁気カセット、フラッシュメモリカード、デジタルビデオディスク、ランダムアクセスメモリ（RAM）、リードオンリメモリ（ROM）などのように、コンピュータがアクセスできるデータを格納しておくことができる、他のタイプのコンピュータ読取可能媒体を、例示の動作環境で使用することも可能である。

【0019】いくつかのプログラムモジュールは、ハードディスク144、磁気ディスク148、光ディスク152、ROM 138またはRAM 140に格納しておくこと

が可能であり、その中には、オペレーティングシステム158、1または2以上のアプリケーションプログラム160、他のプログラムモジュール162、およびプログラムデータ164が含まれている。ユーザは、キーボード166やポインティングデバイス168などの入力デバイスを通して、コマンドや情報をコンピュータ130に入力することができる。他の入力デバイス（図示せず）として可能なものとして、マイクロホン、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナ、1つまたは2つ以上のカメラなどがある。上記および他の入力デバイスは、バス136に結合されたインタフェース170を通して処理ユニット132に接続されている。モニタ172または他のタイプのディスプレイデバイスも、ビデオアダプタ174などのインタフェースを介してバス136に接続されている。モニタのほかに、パーソナルコンピュータは、スピーカやプリンタのような、他の周辺出力デバイス（図示せず）を装備しているのが代表的である。

【0020】コンピュータ130は、リモートコンピュータ176のような、1つまたは2つ以上のリモートコンピュータとの論理コネクションを使用するネットワーク環境で動作するのが通常である。リモートコンピュータ176としては、別のパーソナルコンピュータ、サーバ、ルータ、ネットワークPC、ピアデバイスまたは他の共通ネットワークノードにすることが可能であり、図1にはメモリストレージデバイス178だけが示されているが、コンピュータ130に関連して上述したエレメントの多くまたはすべてを装備しているのが代表的である。図1に示す論理コネクションとしては、ローカルエリアネットワーク（local area network - LAN）180と広域ネットワーク（wide area network - WAN）182がある。この種のネットワーク環境は、オフィス、企業内コンピュータネットワーク、イントラネット、およびインターネット（the Internet）ではありふれた存在になっている。

【0021】LANネットワーク環境で使用されるときは、コンピュータ130は、ネットワークインタフェースまたはアダプタ184を通してローカルネットワーク180に接続されている。WANネットワーク環境で使用されるときは、コンピュータ130は、インターネットのような、広域ネットワーク182上のコミュニケーションを確立するために、モデム186またはネットワークインタフェースのような他の手段を装備しているのが代表的である。モデム186は内蔵型と外付け型があり、どちらも、シリアルポートインタフェース156を介してバス136に接続されている。ネットワーク環境では、パーソナルコンピュータ130に関連して上述したプログラムモジュールまたはその一部は、リモートメモリストレージデバイスに置いておくことが可能である。当然に理解されるように、図示のネットワー

10

20

30

40

50

クコネクションは例示であり、コンピュータ間の通信リンクを確立する他の手段を使用することも可能である。

【0022】一般的に、コンピュータ130のデータプロセッサは、コンピュータの種々コンピュータ読取可能媒体に、それぞれ時間を異にして格納されている命令によってプログラムされている。プログラムとオペレーティングシステムは、例えば、フロッピディスクやCD-ROM上に分散化されているのが代表的である。プログラムとオペレーティングシステムは、フロッピディスクやCD-ROMからコンピュータの二次メモリにインストールまたはロードされる。実行時には、プログラムとオペレーティングシステムは、少なくともその一部が、コンピュータの一次エレクトロニックメモリにロードされる。ここで説明している本発明によれば、マイクロプロセッサまたは他のデータプロセッサに関連して以下に説明されているステップを実現するための命令またはプログラムは、上述したタイプまたは他の種々タイプのコンピュータ読取可能記憶媒体に格納されている。また、本発明によれば、コンピュータ自体は、以下に説明されている方法および手法に従ってプログラムされている。

【0023】説明の便宜上、オペレーティングシステムのように、プログラムや他の実行可能プログラムコンポーネントは、ここでは離散的ブロックとして示されているが、当然に理解されるように、これらのプログラムとコンポーネントは、時期を異にしてコンピュータの異種ストレージコンポーネントに置かれていて、コンピュータのデータプロセッサによって実行される。

#### 【0024】顔面の構造と特性を取り込むための例示システム

従来は、取り込みシステム(capturing system)は、顔全体の顔構造と反射率特性の両方を、相互から独立して同時に取り込むことができなかった。例えば、構造化光(structured light)を使用して顔の構造を取り込むシステムが存在しているが、これらのシステムは、反射率(reflectance)のような、顔の特性を取り込んでいない。同様に、顔の反射率を取り込むシステムが存在しているが、これらのシステムは、顔の構造を取り込んでいない。顔の構造と反射率を、相互に独立して同時に取り込むことができると、収集したデータに別の操作を加え、さまざまな顔レンダリングとアニメーション化操作に役立てることが可能になる。以下では、例示レンダリング操作の一具体例が説明されているが、当然に理解されるように、下述しているシステムと方法の結果として得られる情報またはデータは、他のさまざまな分野で利用することが可能である。例えば、アプリケーション分野としては、セキュリティやパーソナルユーザ間のやりとり、などを目的とした顔の認識、ゲーム、ムービ(映画、動画)などのアニメーションを目的としたリアル感のある顔モデルの構築、およびユーザが自分の顔を容易に取り込んで、インタラクティブエンターテイメントや

ビジネスコミュニケーションで使用できるようにすること、などがある。

【0025】図2は、被写体の顔の顔構造と反射率特性を同時に取り込む上で、使用するのに適している例示システム200を示す。この例示システムは、1つまたは2つ以上のカメラの形体をしたデータ取り込みシステムを装備し、その1つがカメラ202として例示されている。カメラ202は、CCD画像センサと関連回路を装備し、CCDアレイを動作させ、CCDアレイから画像を読み取り、画像をデジタル形体に変換し、その画像をコンピュータに通知することを可能にしている。また、システムは、複数の光源またはプロジェクタの形体をした顔照明システム(facial illumination system)も装備している。複数のカメラが使用される場合には、これらのカメラは、時間的に同時取り込みを可能にするようにジェンロック(genlock)されている。図示の例では、2つの光源204、206が利用されている。光源204は、望ましくは、被写体の顔の上に投影される構造化パターン(structured pattern)を出力する。光源204は、適当なロケーションであれば、どこにでも置くことが可能になっている。このパターンは、被写体の顔の3D形状に関する構造情報またはデータを、カメラ202で取り込むことを可能にしている。赤外線領域の光で構成されたパターンを採用すると有利であるが、適当な光源であれば、どの光源でも使用可能である。光源206は、望ましくは、顔の反射率特性の拡散反射成分をカメラ202で取り込めるようにする光を出力する。光源206は、例えば、ビームスプリット(beam splitting)手法でカメラのレンズ202aと同列に置いておくことが有利であるが、適当なロケーションであれば、どこにでも置くことが可能である。この光源は、カメラのレンズを取り囲むように適応させることも可能である。また、この光源は、反射率の鏡面反射成分(specular component)が抑止または除去されるように選択されている。図示の例では、線形偏光フィルタ(linear polarizing filter)は偏光照明を得るために採用され、別の偏光フィルタは前記線形偏光フィルタに直交する向きになっていて、レンズ202aの前に置かれているため、顔からの鏡面反射がカメラによって記録されないようにしている。上述した照明システムは、異なる周波数の光源、例えば、カメラの赤と緑チャネルに対応する光源を使用してシミュレートされている。なお、これらのチャネルは、どちらも赤外線領域にすることが可能である。さらに、光源がナローバンド(例えば、780-880 nm)になるように選択すると、周囲光の影響を除去することができる。この特性が達成されるのは、カメラもナローバンドになるようにフィルタリングされるときだけである。光源からの照明はナローバンドの波長に集中されるのに対し、周囲光はブロードレンジの波長にわたって拡散されるので、光源からの光は、上記のような波長のときの周囲光よりも

強力になる。カメラは、光源から放出された波長だけを記録するようにフィルタリングされるので、周囲光による影響を受けることはほとんどない。その結果、カメラは、選択された光源が被写体に及ぼす影響だけを検出することになる。

【0026】複数の異なる光源を使用すると、具体的には、赤外線光源を偏光光源（これも赤外線光源にすることが可能である）と併用すると、カメラは、顔に関する構造情報またはデータ（光源204からの）と顔に関する反射率情報またはデータ（光源206からの）を、相互に独立して同時に取り込むことが可能になる。構造情報は顔の3次元様相を記述しているのに対し、反射率情報は顔の拡散反射率特性を記述している。この情報は、コンピュータ208などの、コンピュータ化画像プロセッサによって処理されると、以後の顔アニメーション操作で利用できる情報またはデータが得られることになる。以下で説明する例では、この情報は3次元データ（3Dデータ）とアルベドマップ（albedo map）を含んでいる。

【0027】図3は、ここで説明する実施形態による方法のステップを記載しているフローチャートである。ここで説明する方法によれば、顔の構造と反射率特性に関する情報またはデータは、同時に収集し、処理することが可能になっている。ステップ300で、被写体の顔は複数の異なる光源で照射される。このステップを実現する例示システムは図2に示されている。なお、図示の例では、2つの例示光源が利用されているが、当然に理解されるように、他の複数の光源を使用することも可能である。ステップ302で、ステップ300の照射からレンジマップデータ（デプスマップデータ）と画像データが測定される。すなわち、ステップ300の照射により、カメラは光反射率を検出することが可能になり、この光反射率は、レンジマップデータと、そこに構造光源パターンを含んでいない画像データ（つまり、反射率）の両方を得るために利用されることになる。レンジマップデータと画像データはコンピュータ208（図2）に渡され、そこで処理される。この時点で、ステップ304で、汎用顔テンプレート（generic face template）をレンジマップデータに適用すると、レンジマップデータに関連して生じている種々のノイズを除去することができるが、これはオプションである。汎用顔テンプレートは、レンジマップデータに含まれるノイズを除去する3Dフィルタと考えることができる。汎用顔テンプレートは、この分野の精通者に自明のものである。

【0028】ステップ306で、レンジマップデータが使用され、3D面が導き出され、あるいは計算される。適当なアルゴリズムであれば、どのアルゴリズムでも使用可能であり、これは、この分野の精通者に自明のものである。アルゴリズムの例としては、次の論文に記載されているものがある。すなわち、Turk & Levoy著「レンジ

画像からのジップアドポリゴンメッシュ（Zippered Polygon Meshes）」（SIGGRAPH94）、およびF. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, G. Taubin共著「面再構築のためのボールピボティングアルゴリズム（The Ball-Pivoting Algorithm）」（Trans. Vis. Comp. Graph. 5:4（1999））である。ステップ308で、ステップ306の3D面に対する面法線ベクトル（「面法線」）が、公知のアルゴリズムを使用して計算される。このタスクを実行する1つの方法は、三角形に対する法線を計算し、各頂点周りの三角形法線の平均をとって、頂点法線を作り、頂点法線を各三角形の内部を横切るように補間することである。当然のことであるが、他の方法を利用することも可能である。次に、ステップ310で、ステップ308で計算された面法線とステップ302の画像データが使用されて、アルベドマップが導き出される。アルベドとは、特殊タイプのテクスチャマップであり、そこでは、各サンプルは、顔面上の特定ポイントでの顔面の拡散反射率を記述している。アルベドマップの導出は、上記に示した情報が与えられていれば、この分野の精通者には自明のものである。例示のアルゴリズムは、Marschner著「コンピュータグラフィックスの逆レンダリング（Inverse Rendering for Computer Graphics）」（PhD thesis, Cornell University, August 1998）に記載されている。

【0029】この時点で、図2に示すように、照明処理の結果として、被写体の顔の構造特徴を記述している3Dデータと、画面の拡散反射率を記述しているアルベドマップデータとが得られている。

【0030】上記照明処理は、記述されている情報を抽出するために使用できるので、抽出した情報は適当な目的のために使用することができる。特に利点のある実施形態では、抽出された情報は、被写体の表情を抽出し、認識するために利用されている。この情報は、表情変換のために使用できる。すぐあとで説明されている、本発明による実施形態では、ある人物の表情は、別人物の表情を、リアル感をもって変換するために使用することができる。

【0031】コードブックの使用による表情変換ある表情変換実施形態では、コードブック（code book）の考え方が取り入れられ、以下に説明されている表情変換操作で利用されている。図4は、ある人物から取り込まれた多数の、異なる表情を収めているコードブック400の例を示す。これらの表情は、汎用表情（generic expressions）、つまり、特定の個人からではなく、汎用的人物からの表情と考えることができる。この例では、表情は、表情1から表情Nまでの範囲を有する。表情1は、例えば、「スマイル」であり、表情2は「しめ面」にであり、表情3は「怒り」の表情である。以下、同様である。コードブック400に収められている表情は、数学的には、それらのジオメトリの形で記述さ

れており、すぐ上で説明したプロセスのような、適当な方法で取り込むことが可能である。

【0032】表情変換を行うために、コードブック400内の表情の一部を使用して、変換関数(transformation function)が最初に導き出される。変換関数を導き出すために、トレーニング表情集合(training set of expressions)402という考え方が取り入れられている。表情トレーニング集合402は、その表情がコードブック400に記述されている個人とは別の個人から得られる表情の集合で構成されている。トレーニング集合402のトレーニング表情は、コードブック表情のサブセット(部分集合)になっている。つまり、トレーニング集合の中の各表情は、コードブック400の中の表情に対応している。例えば、トレーニング集合402は、3つの表情、すなわち、表情1、表情2、および表情3で構成することが可能であり、そこでは、表情は、それぞれ「スマイル」、「しかめ面」および「怒り」になっている。変換関数の目標は、トレーニング集合の表情に関連付けられたジオメトリック変形を受け取り、そのジオメトリック変形をコードブック400の表情のすべてに適用して、コードブック表情がリアル感のある表情表現となるようにすることである。すなわち、各人物の顔は、どの任意の表情についても、ジオメトリが異なった変形の仕方をするものと考えられる。ある表情のときの、ある人物のジオメトリックな顔の変形が、表情をレンダリングする目的で別人物の顔に単純に適用されると、その変形が適用された顔は、非常にゆがんだように見えることになる。このようなことが起こるのは、顔のジオメトリが異なるだけでなく、顔と顔の間のように顔の変形が異なるためである。従って、ある表情集合から別の表情集合への最良変換が得られるような変換関数が導き出されている。

【0033】図4に戻って説明すると、図4には、線形変換プロセッサ406が示されている。この変換プロセッサ406は、適当なハードウェア、ソフトウェア、ファームウェア、あるいはこれらの組み合わせで実現することが可能である。図示の例では、線形変換プロセッサ406はソフトウェアで実現されている。線形変換プロセッサはトレーニング表情集合402と、対応するコードブック表情404とを入力として受け取る。変換プロセッサはこれらの入力を処理して、変換関数408を導き出す。この変換関数408を、コードブック400内の表情のすべてに適用すると、合成表情集合(synthesized set of expressions)が得られる。この合成表情集合は、トレーニング表情集合が得られた人物の表情に関連するジオメトリック変形によって操作された、コードブックの表情を表している。

【0034】表情が同じであるときの顔の変位は、次の2つの理由で異なる人物では同じになっていない。その1つは、動き取り込みサンプルポイント(この具体例で

は、この特定アルゴリズムによると、顔の動きがどのように表現できるかを示している)は、位置に誤差があると、正確に一致しないからである。もう1つは、頭の形状とサイズは人物と人物の間で変化するからである。

【0035】第1の不一致は、すべての顔の動き取り込み変位データを、汎用メッシュ上の一定位置集合で再サンプリングすることによって解消することができる。これについては、以下の「顔モデルを構築するための例示システムと方法」の個所に詳しく説明されている。そこでは、一定位置集合は「標準サンプル位置(standard sample positions)」と呼ばれている。再サンプリング関数はメッシュ変形関数(mesh deformation function)である。標準サンプル位置は、一度細分化された汎用メッシュの頂点に対応する、顔メッシュの頂点である。

【0036】第2の不一致では、ある顔から別の顔へ変位データを変換し、顔のサイズと形状の変化を補償する必要がある。図示の例では、これは、2つのデータセットについて対応する表情の小さなトレーニング集合を見付けたあと、一方から他方への最良線形変換を見付けることによって行われる。1つの例として、次の場合を考えてみる。すなわち、実験的環境では、感情を表す表情は、複数の表情のさまざまな強さを含む、49個の対応する表情に対して手作業でラベルが付けられていた。音声の動きについては、10,000フレームは、時間歪曲(time warping)手法を用いて自動的にアライメントされていた。

【0037】各表情は、 $m$ 個の標準サンプル位置での $x$ ,  $y$ ,  $z$ 変位のすべてを含んでいる $3m$ ベクトル $g$ によって表されている。変換される顔の $n$ 個の表情ベクトル集合を $g_{a1...n}$ とし、ターゲット顔の対応するベクトル集合を $g_{b1...n}$ とすると、線形予測子(linear predictors)集合 $a_j$ は、下に示す $3m$ 線形最小二乗システムの解を求めることによって、 $g_n$ の座標ごとに1つずつ計算される。

【0038】

【数3】

$$a_j \cdot g_{ai} = g_{bi} \quad i=1...n$$

【0039】図示の例では、各 $g_{aj}$ のポイントは小さなサブセットだけが使用されている。具体的には、標準サンプルポイントとエッジを共有するポイントだけが考慮の対象になっている。使用されたメッシュでは、平均価(average valence)は約6であるので、代表的な $g_{aj}$ は18要素になっている。その結果のシステムは大体 $n \times 18$ である。

【0040】結果の線形システムは条件付けが十分でない場合があり、その場合には、線形予測子 $a_j$ は良好に概念化されない。特異値(singular value)の拡散は、擬似逆関数(pseudoinverse)を計算して、 $a_j$ の解を求めるとき制御されるので、概念化が大幅に改善される。 $\alpha$   $\sigma_1$ 未満の特異値はすべてゼロにされる。ただし、 $\sigma_1$ は

マトリックスの最大特異値であり、 $\alpha = 0.2, \dots, 0.1$ である。

【0041】図5は、ここで説明している実施形態による表情変換方法のステップを記載しているフローチャートである。ステップ500で、表情のコードブックが用意される。この種のコードブックの例は上に示されている。ステップ502で、トレーニング表情集合が用意される。代表例として、このトレーニング集合は、コードブック表情が得られた人物とは異なる人物からの表情の集合になっている。トレーニング表情集合は、適切であれば、どの方法によっても取り込むことができる。1つの例として、表情は、図2に示すようなシステムを使用して取り込むことができる。トレーニング表情集合が用意されたあと、ステップ504で、トレーニング集合とコードブックを使用して変換関数が導き出される。この作業を行う1つの方法例は、上述した通りである。当然のことであるが、本発明の精神と範囲を逸脱しない限り、他の方法を使用することも可能である。例えば、ニューラルネットワーク (neural network) などの、種々の非線形変換を使用することも、基礎表情 (basis expressions) の加重和を使用することも可能である。変換関数が導き出されると、これはコードブック内の表情のすべてに適用され、以後の顔アニメーション操作の基礎として利用できる合成表情集合が得られ、あるいは定義される。

#### 【0042】例示のアプリケーション

図6はシステム600を示す。そこには、上述した表情変換プロセスをどのように採り入れるかが、1つの例で示されている。システム600は送信側コンピューティングシステム、つまり、トランスミッタ602と、受信側コンピューティングシステム、つまり、レシーバ604とを含み、これらは、インターネット (the Internet) などのネットワーク603を介して通信するように接続されている。トランスミッタ602は照明システム200 (図2) を含み、これは、図2を参照して上述したように、ある人物の表情を取り込むように構成されている。さらに、トランスミッタ602は、図4を参照して上述したものと同一ような、コードブック400を含んでいる。なお、ここでは、コードブックは、上述したように合成表情集合に合成されているものとする。すなわち、照明システム200が取り込んだ表情をもつ人物から得られたトレーニング表情集合を使用して、コードブックが処理され、合成表情集合が得られている。

【0043】レシーバ604は再構築モジュール606を含み、これは、トランスミッタ602から受信したデータから顔画像を再構築するように構成されている。また、レシーバ604は、トランスミッタ602側のそれと同じコードブック400を含んでいる。ここでは、トランスミッタ602側にいる人物が、レシーバ604側にいる人物と通信しようとしている。トランスミッタ6

02側にいる人物が通信のために顔を動かすと、その顔表情と動きはトランスミッタ602によって取り込まれ、処理される。この処理には、顔表情を取り込み、合成コードブックをサーチして、コードブックの中で最も近似している合致表情を見付けることを含めることが可能である。合致する表情が合成コードブックの中で見つかり、その表情のインデックスがレシーバ604に送られ、アニメーション化された顔が再構築モジュール606を使用して再構築される。

#### 10 【0044】例示の顔変換

図7は、ここで説明している実施形態による表情変換のいくつかの効果を示す図である。1行目の絵 (A) は、第1人物 (人物A) の合成顔を構成しており、そこには3つの異なる表情が示されている。これらの絵は、人物Aの取り込まれた顔の動きの結果である。第2人物 (人物B) の顔の動きは、すでに取り込まれている。人物Bの取り込まれた顔の動きは3行目 (C) に示されている。ここでは、3D動きデータは、いくつかのカラードットを人物の顔の上に置き、人物の顔が変形されたときドットの動きを測定することによって取り込まれているが、このことは、この分野の精通者ならば当然に理解されるものである。なお、顔の動きは、上述したシステムと方法によって取り込むことができる。人物Bの取り込まれた動きは、上述したように使用されて、人物Aの表情が変換されている。この操作の結果は、2行目に示されている。これら3セットの絵の表情は、すべて相互に一致している。同図に示すように、1行目 (A) と2行目

(B) の表情は、非常に異なる2人物から導き出されたとしても、非常に類似しているのに対し、第2人物 (3行目 (C)) のオリジナル表情は、1行目 (A) と2行目 (B) の表情とは全体的に類似していない。

#### 30 【0045】顔モデルを構築するための例示システムと方法

リアル感のある画像を作るために必要とされる顔のモデルは、2つの部分からなっている。モデルの第1部分は、顔のジオメトリ (つまり、顔面の形状) に関係している。他方、モデルの第2部分は、顔の反射率 (つまり、顔のカラーと反射特性) に関係している。このセクションでは、このモデルの第1部分、つまり、顔のジオメトリについて説明する。

40 【0046】顔のジオメトリは、皮膚面に、目を表す別の面を加えたものから構成されている。この例では、皮膚面は、頭をレーザレンジでスキャンすることによって導き出され、変位マップをもつ細分面で表されている。目は別モデルであり、これは皮膚面と位置合わせされ、マージされて、高品質レンダリングに適した完全顔モデルが得られている。

#### 【0047】メッシュ適合

顔モデルを構築する第1ステップでは、レーザレンジスキャナで測定されたジオメトリに密接に近似する細分面

が作成される。図示の例では、細分面は、Loopの細分ルールを使用して粗三角メッシュから定義されている。Loopの細分ルールは、Charles Loop著「三角形に基づく平滑細分面(Smooth Subdivision Surfaces Based on Triangles)」(PhD thesis, University of Utah, August 1987)に詳しく説明されている。さらに、細分面は、Hoppe他著「区分平滑面再構築(Piecewise smooth surface reconstruction)」(Computer Graphics (SIGGRAPH '94 Proceedings) pp. 295-302, July 1994)に記載されているものと類似のシャープエッジを追加して定義されている。なお、不規則なクリーズマスク (crease mask)は使用されていない。さらに、ダーツ(dart)とクリーズ頂点(crease vertex)の間のエッジを細分化するとき、クリーズ頂点に隣接する新エッジだけがシャープエッジとしてマークされている。

【0048】すべての顔モデルに対する細分面は、シングルベースメッシュを使用して定義され、頂点位置だけが、各々の異なる顔の形状に適応するように変化している。図示の例では、227個の頂点と416個の三角形をもつベースメッシュは、顔の全体形状をもち、目と唇付近が詳細化されるように定義されており、そこでは、最も複雑なジオメトリと動きが現れている。口の開きはメッシュの境界になっており、上唇と下唇上の対応する頂点の位置を1つに結ぶことによって、適合プロセス期間中は、閉じたままになっている。このベースメッシュには、シャープな細分ルール用にマークされた少数のエッジがあり、これらは、口の開きの両側にコーナを作成し、鼻の両側を折り重ねるときの場所を得るのに役立っている。修正された細分ルールでは、少なくとも3つのシャープエッジのチェーンに対してだけクリーズが取り入れられているので、このモデルは表面にクリーズがない。孤立頂点(isolated vertex)だけは、明確に定義された限界法線(limit normal)をもつことができないからである。

【0049】図8は、この例に従って使用された粗定義メッシュ(中央の図)の例を示す。図8は、この粗メッシュをどのように使用して、同じ細分制御(粗)メッシュを各々の顔の変位細分面にマッピングすると、ある顔と別の顔との間に自然な対応関係が得られるかを示したものである。この側面は、以下に詳しく説明されている。

【0050】細分面を各顔に適合させるために使用されるプロセスは、Hoppe他著「区分平滑面再構築(Piecewise smooth surface reconstruction)」(Computer Graphics (SIGGRAPH '94 Proceedings) pp. 295-302, July 1994)に記載されているアルゴリズムに基づいている。Hoppeの面適合法は、基本的には、3つのフェーズからなるものと言うことができる。すなわち、トポロジ型推定(フェーズ1)、メッシュ最適化(フェーズ2)、および区分平滑面最適化(フェーズ3)である。

【0051】フェーズ1では、未知の面上または付近に未組織のポイント集合が与えられている、相対的に多数の三角形からなる三角メッシュが構築される。このフェーズでは、面のトポロジ型が判断され、ジオメトリの初期推定が得られる。フェーズ2は、フェーズ1の出力から始まり、三角形の数を減らし、データへの適合を改善する。このアプローチでは、表現の簡潔化と適合の良好化という、相互に競合する目標間のトレードオフを明示的にモデル化するエネルギー関数を最適化することを課題としている。この最適化手順における自由変数としては、メッシュ内の頂点の数、その接続性(connectivity)、およびその位置がある。フェーズ3は、フェーズ2で得られた最適化メッシュ(区分線形面)から出発し、この場合も、簡潔性とデータへの適合性をトレードオフするエネルギー関数を最適化することによって、正確で、簡潔な区分平滑細分面を適合させる。フェーズ3で最適化が行われると、制御メッシュ内の頂点の数、その接続性、その位置、およびシャープな特徴の数とロケーションが変化する。面におけるシャープな特徴の自動検出と回復は、このフェーズの不可欠な部分となっている。

【0052】この実施形態では、処理の仕方が、Hoppe他に記載されているアプローチとは、2つの点で異なっている。第1は、制御メッシュの接続性を変更する必要がないので、連続的最適化は、頂点位置に対してだけ行われることである。さらに、特徴の制約が、平滑化項と共に追加されていることである。

【0053】図示の例では、適合プロセスは次の汎関数を最小化する。

【0054】

【数4】

$$E(\mathbf{v}) = E_d(\mathbf{v}, \mathbf{p}) + \lambda E_s(\mathbf{v}) + \mu E_c(\mathbf{v})$$

【0055】上記において、 $\mathbf{v}$ はすべての頂点位置のベクトル、 $\mathbf{p}$ はレンジスキャナからの全データポイントのベクトルである。3つの項の下付き文字は、距離、形状、および制約を表している。距離関数 $E_d$ は、レンジスキャナポイントから細分面までの和二乗距離(sum-squared distance)を測定する。

【0056】

【数5】

$$E_d(\mathbf{v}, \mathbf{p}) = \sum_{i=1}^{n_p} a_i \|\mathbf{p}_i - \Pi(\mathbf{v}, \mathbf{p}_i)\|^2$$

【0057】上記において、 $\mathbf{p}_i$ はI番目のレンジポイント、 $\Pi(\mathbf{v}, \mathbf{p}_i)$ は、頂点位置 $\mathbf{v}$ で定義された細分面上のレンジポイントの投影である。重み $a_i$ は、 $\mathbf{p}_i$ におけるスキャナの視点方向が $\Pi(\mathbf{v}, \mathbf{p}_i)$ における面法線と一致していないポイントを無視させるブール項である。さらに、面からある距離以上に離れているポイントは拒否される。

【0058】

【数 6】

$$a_i = \begin{cases} 1 & \text{if } \langle s(p_i), n(\Pi(v, p_i)) \rangle > 0 \text{ and } \|p_i - \Pi(v, p_i)\| < d_0 \\ 0 & \text{otherwise} \end{cases}$$

【0059】上記において、 $s(p)$  は、ポイント  $p$  におけるスキヤナの視点に向かう方向、 $n(x)$  はポイント  $x$  における外側に面する面法線である。

【0060】平滑関数  $E_s$  は、制御メッシュをローカルに平坦化させるものである。この関数は、各頂点から隣接頂点の平均までの距離を測定する。

【0061】

【数 7】

$$E_s(v) = \sum_{j=1}^{n_v} \left\| v_j - \frac{1}{\deg(v_j)} \sum_{i=1}^{\deg(v_j)} v_{k_i} \right\|^2$$

【0062】頂点  $v_{k_i}$  は、 $v_j$  の隣接頂点である。

【0063】制約関数  $E_c$  は、単純には、制約頂点集合から対応するターゲット位置集合までの和二乗距離である。

【0064】

【数 8】

$$E_c(v) = \sum_{i=1}^n \|A_i v - d_i\|^2$$

【0065】上記において、 $A_j$  は、 $j$  番目の頂点の限界位置を、制御メッシュからとらえて定義し、頂点  $c_i$  の限界位置が 3D ポイント  $d_i$  に付加されるようにする線形関数である。制約は、最適化変数の再パラメータ化によって厳格に適用することも可能であるが、この制約をソフトにすると、繰り返しが望ましいローカル最小限まで円滑に行われることが判明している。制約は、汎用メッシュの顔特徴を、適合される特定の顔上の、対応する特徴に合致させるためにユーザによって選択される。この例では、約 25 乃至 30 の制約が使用され、目、鼻、および口に集中している。図 9 は、900 の細分制御メッシュ (A) の制約と、顔モデル (B) 上の対応するポイントを示す。

【0066】 $E(v)$  を最小限にすることは、 $\Pi$  と  $a_i$  が  $v$  の線形関数でないので、非線形最小二乗問題である。しかし、 $a_i$  を一定に保ち、 $\Pi(v, p_i)$  を一定の線形的制御頂点の組み合わせによって近似化すると、これを線形問題にすることができる。従って、適合プロセスは、線形最小二乗問題のシーケンスとして進められ、 $a_i$  および面上の  $p_i$  の投影は、各繰り返しの前に再計算されることになる。細分境界面は、これらの計算を行うために、特定の細分レベルにあるメッシュによって近似化される。顔を適合することは、少数の繰り返しで行われ (30 未満)、制約は、繰り返しの進行と共に、単純なスケジュールに従って更新されて行く。その更新は高  $\lambda$  と低  $\mu$  が

ら始まって、最適化を非常にスムーズな顔の近似化に導いて行き、低  $\lambda$  と高  $\mu$  まで進んで行き、最終的解により、データと制約は密接に適合することになる。実際の計算時間は、 $\Pi(v, p_i)$  の計算時間によって決まる。

【0067】レンダリングのためのメッシュを作るために、面は望ましいレベルまで細分化され、顔の形状にスムーズに近似しているメッシュが得られる。変位は、各々の頂点ごとに、その頂点での面に対するライン法線を、オリジナルスキャンで定義された三角面と交差させることによって計算される。これについては、Lee 他著「変位細分面 (Displaced Subdivision Surfaces)」(SIGGRAPH '00 Proceedings, July 2000) に説明されている。その結果の面は、オリジナルスキャンの目立った特徴をすべて、若干三角形が少なくなっているメッシュに再現しているが、これは、ベースメッシュは顔の重要な領域の三角形が多くなっているためである。この細分ベースの表現によると、面をパラメータ化したものと、そのパラメータ化で定義された組み込み多重解像度基底関数集合 (built-in set of multiresolution basis functions) とが得られることになり、さらに、特徴制約が適合時使用されるため、この方法を使用して適合されるすべての顔にわたって自然の対応関係が得られることになる。この構造は、顔のアニメーションで利用すると、多くの面で便利である。

【0068】図 10 は、ここで説明している実施形態に従って顔モデルを構築するための方法のステップを記載しているフローチャートである。この方法は、適当なハードウェア、ソフトウェア、ファームウェアまたはこれらの組み合わせで実現することが可能である。この例では、方法はソフトウェアで実現されている。

【0069】ステップ 1000 で、1 つまたは 2 つ以上の顔の 3D データが測定され、対応する顔モデルが得られる。上記の例では、3D データは、顔をレーザレンジでスキャンすることによって得られていた。当然に理解されるように、3D データは他の方法で得ることも可能である。ステップ 1002 で、1 つまたは 2 つ以上の顔モデルに適合するために使用される汎用顔モデルが定義される。当然に理解されるように、汎用顔モデルは、多数の異なる顔に適合するように利用できるという利点がある。従って、このことは、そのようなことが行われていなかった従来の方法に比べて、改良点となっている。上述した例では、汎用顔モデルは、粗三角メッシュ (coarse triangle mesh) の形体をしたメッシュ構造になっている。この三角メッシュは、顔のジオメトリに密接に近似している細分面を定義している。図示の例では、顔モデルのすべてについての細分面は、シングルベースメ



シュを使用して定義されている。ステップ1004で、汎用顔モデル上の個々のポイントまたは制約が選択される。これらの個々のポイントまたは制約は、顔モデル上にマークされた、対応するポイントに直接にマッピングされる。これら個々のポイントのマッピングは、起こり得る多数の、異なる顔モデルの各々について、同じように行われる。ステップ1006で、汎用顔モデルは1つまたは2つ以上の顔モデルに適合される。このステップは、各々の異なる顔の形状に適応するように、頂点の位置だけを操作することによって行われる。適合プロセス期間には、頂点位置だけにわたって連続的最適化が行われるので、メッシュの連続性が変更されることはない。さらに、この適合プロセスでは、個々のポイントまたは制約が直接に顔モデルにマッピングされる。さらに、平滑項 (smoothing term) が追加され、これは、制御メッシュがローカルに平坦化されるように最小化されている。

#### 【0070】目の追加

上述してきた変位細分面は、顔の皮膚面の形状を非常によく表している。しかし、リアル感のある顔を得るために望ましいとされる特徴が、他にもいくつかある。その中で最も重要な特徴が目である。レーザレンジスキャナは目に関する適切な情報を取り込まないので、メッシュは、別々にモデル化された目を追加することによって、レンダリングのために補強されている。顔モデルの他の部分と異なり、目とその動きは、特定の人物から測定されないの、リアルな目の外観を再現するとは限らないようになっている。しかし、目の存在と動きは、顔モデルの全体的外観にとって非常に重要である。

【0071】適切な目モデルであれば、その目モデルを使用して目をモデル化することができる。図示の例では、2部分からなるモデルを構築するために、市販のモデリングパッケージが使用されている。第1部分は眼球のモデルであり、第2部分は目回りの皮膚面のモデルであり、その中には、まぶた、眼窩、および周囲顔面部分が含まれている（この第2部分は「眼窩面 (orbit surface)」と呼ぶことにする）。目を全体顔モデルの一部にするためには、眼窩面を、モデル化しようとする個人の顔に適合させる必要があり、2つの面を縫い合わせる必要がある。これは2ステップで行われる。最初に、2つのメッシュは、眼窩面上に定義された重み付け関数に従ってゆがみが付けられ、顔と眼窩が重なり合った個所で同時に現れるようにする。次に、2つ面はペアの同心楕円面でカットされ、1つのメッシュになるように縫い合わされる。

【0072】なお、上述してきた実施形態の利点のある特徴の1つは、これらの実施形態によると、ある人物の表情を別の人物の表情に変換するために使用できる構造またはフレームワークが得られることである。汎用顔モデルを各個人の顔に適合することは、汎用モデルのどの

部分も、各人物の顔上の同じ特徴に常にマッピングされるように制約されている。例えば、汎用モデル中の口の左隅は、どの人物の顔でも、その口の左隅に常にマッピングされる。適合顔モデルの集合を利用すると、どの顔においても、ある特定の顔にある特定のポイントに対応するポイントを判断することができる。例えば、人物Aの顔上の口の左隅の動きが測定されたとする。顔Aへの汎用モデルの適合を使用すると、汎用モデルのどのポイントが測定ポイントに対応しているかを判断することができ、次に、顔Bへの汎用モデルの適合を使用すると、Bの顔のどのポイントが汎用モデル上の、計算されたポイントに対応しているか、従って、顔Aの測定ポイントに対応しているかも判断することができる。測定した顔上のポイント集合から動きを再現するために新しい顔のどの部分を移動させる必要があるかを知っている必要があるため、この情報は、動きをある顔から別の顔に変換する上で不可欠である。

#### 【0073】顔の動き

顔の動きは、画面上のサンプルポイント集合の、時間と共に変化する3D位置によって指定される。顔が動き取り込みデータによって制御されるとき、これらのポイントは、動き取り込みシステムによってトラッキングされる、顔上のマーカになっている。これらのポイントの動きは、顔面の領域にスムーズに影響する制御ポイントの集合によって顔面を制御するために使用される。顔の動きデータの取り込みは、この分野の精通者ならば理解されるように、いずれかの適当な方法で行うことができる。1つの具体例では、顔の動きは、Guenter他著「顔の作成 (Making Faces)」(Proceedings of SIGGRAPH 1998, pp. 55-67, 1998) に記載されている手法を使用して取り込まれていた。

#### 【0074】メッシュ変形

顔は、制御ポイント集合 $q_j$ の変位の線形的組み合わせに従って、三角メッシュの各頂点 $w_i$ をその静止位置から変位することによってアニメーション化されている。これらの制御ポイントは、動きを記述しているサンプル点 $p_j$ と1対1で対応している。各制御ポイントが頂点に及ぼす影響は、対応するサンプルポイントからの距離と共に低下していき、複数の制御ポイントが頂点に影響する場合には、それらの重みは総和が1になるように正規化される。

#### 【0075】

#### 【数9】

$$\Delta w_i = \frac{1}{\beta_i} \sum_j \alpha_{ij} \Delta q_j : \alpha_{ij} = h(\|w_i - p_j\|/r)$$

【0076】上記において、頂点 $i$ が複数の制御ポイントによって影響される場合、 $\beta_i = \sum_k \alpha_{ik}$  であり、そうでなければ1である。これらの重みは、サンプルポイントと顔メッシュの静止位置を使用して一回計算される

ので、フレームごとのメッシュの動きはスパースマトリックス乗算 (sparse matrix multiplication) になっている。重み付け関数については、次式が使用されている。

【0077】

【数10】

$$h(x) = \frac{1}{2} + \frac{1}{2}\cos(\pi x).$$

【0078】顔をアニメーション化する細部を処理するには、これらの重み付けルールには2つタイプの例外が設けられている。目と口の付近の頂点と制御ポイントは、「上」および「下」のタグが付けられ、例えば、口の上にある制御ポイントは口の下にある頂点の動きには影響しない。また、目の周囲の領域におけるスカラーテクスチャマップ (scalar texture map) は、目がスムーズなテーパー状になって、まぶたの個所でゼロになるように動きに重み付けするために使用されている。サンプルポイント集合に従って顔メッシュを動かすには、制御ポイント位置を計算して、顔面がそれなりに変形されるようにしなければならない。上述したのと同じ重み付け関数を使用すると、サンプルポイントが制御ポイントに応じてどのような動きをするかを計算することができる。その結果は線形変換  $p = Aq$  である。従って、時刻  $t$  に、サンプル位置  $p_t$  を求めるには、制御ポイント  $q_t = A^{-1}p_t$  を使用することができる。しかし、マトリックス  $A$  は条件付けが十分でないことがあるので、制御ポイントの動きが非常に大きいために起こる、望ましくない面形状を回避するために、SVD (Singular Value Decomposition—特異値分解) を使用して  $A^{-1}$  が計算され、 $A^{-1}$  の特異値は限界  $M$  でクランプされる。図示の例では、 $M = 1.5$  が使用されている。SVDを論じている参考文献としては、Golub および Van Loan 著「マトリックス計算 (Matrix Computations)」第3版、Johns Hopkins press, 1996がある。

【0079】目と頭の動き

顔を実物に近い外観にするためには、プロシージャで生成された動きが目に加えられ、別々に取り込まれた剛体運動が顔全体に加えられる。眼球は、ランダムな固定方向のシーケンスに従って回転され、一方の眼球から次の眼球に向かってスムーズに移動して行く。まぶたは、眼球の中心を通る軸周りにまぶたを定義している頂点を回転し、まぶたメッシュ上に定義された重みを使用して変形がスムーズに行われるようにすることでアニメーション化される。

【0080】頭の剛体運動は人物の頭の物理的動きから取り込まれ、これは、その人物が特殊なマシン認知可能ターゲットでマークされた帽子を着用している間にその動きをフィルムに写すことによって行われる (帽子は、Marschner他著「人間の皮膚を含む画像ベースBRDF測定 (Image-based BRDF measurements including human skin)」、Rendering Techniques '99 (Proceedings of the

Eurographics Workshop on Rendering), pp. 131-144, June 1999で使用されているものに近似するようにパターン化されている)。これらのターゲットをビデオシーケンスでトラッキングすることにより、頭の剛体運動が計算され、そのあと、これは顔モデルをレンダリングするために適用される。このセットアップによると、ビデオカメラだけで済むので、望みのアクションを実演しながら頭の動きを作っていくことができる。

【0081】反射率をモデリングするための例示システムと方法

リアル感のある顔画像をレンダリングするためには、ジオメトリが正確であるだけでなく、皮膚からの光の反射の計算が正確であることが要求される。図示の例では、顔をレンダリングするために、物理ベースのMonte Carlo光線トレーサが使用されている。手法の例としては、Cook他著「分布光線トレーシング (Distribution Ray Tracing)」、Computer Graphics (SIGGRAPH '84 Proceedings), pp. 165-174, July 1984およびShirley他著「直接照明計算のためのMonte Carlo手法 (Monte Carlo techniques for direct lighting calculations)」、Transactions on Graphics, 15(1):1-36, 1996に記載されているものがある。この方法によると、任意のBRDF (bi-directional reflectance distribution functions: 双方向反射率分布関数) を使用して、単純なシェーディングモデルでは近似化が困難であった、皮膚の外観を正確にシミュレートすることができる。さらに、拡張光源が使用されているが、これは、ポートレート写真のようにレンダリングするとき、見苦しくない画像を得るために必要になるものである。物理的光移動は、計算効率の必要上、次の2点でずれるようにしている。すなわち、拡散相互反射 (diffuse interreflection) は無視され、目は屈折なしで瞳を通して照明されている。

【0082】図示の例では、皮膚の反射率モデルは、実際の人間の顔の測定に基づいている。手法の例としては、Marschner他著「人間の皮膚を含む画像ベースBRDF測定 (Image-based BRDF measurements including human skin)」、Rendering Techniques '99 (Proceedings of the Eurographics Workshop on Rendering), pp. 131-144, June 1999に記載されているものがある。これらの測定は、いくつかの被写体の顔の平均BRDFを記述しており、Lafortune他著「反射率関数の非線形近似化 (Non-linear approximation of reflectance functions)」、Computer Graphics (SIGGRAPH '97 Proceedings), pp. 117-126, August 1997に記載されているBRDFモデルの適合パラメータを含んでいる。従って、これらの測定は、リアル感のある顔をレンダリングするときの、すぐれた出発点となっている。しかし、これらの測定は、実際の顔で観察される空間的変化の一部を含むように補強する必要がある。これは、その皮膚がレンダリングした皮膚と類似している被写体の測定BRDFへの適合から始め、それ

を拡散反射成分と鏡面反射成分に分離することによって行われる。そのあと、各々をモジュレートするためにテクスチャマップが取り入れられている。

【0083】拡散反射成分のテクスチャマップ、つまり、「アルベドマップ(albedo map)」は、以下で説明するように、被写体の実際の顔から取られた測定に従って、拡散反射率をモジュレートする。鏡面反射成分は、スカルテクスチャマップ(scalartexture map)によってモジュレートされ、皮膚反射率でレンダリングしてはならない箇所(まゆやまつ毛など)から鏡面反射が除去され、顔の下部の鏡面反射が低減され、顔の皮膚の特徴に近似するようにされる。その結果は、空間的に変化するBRDFであり、これは、Lafortune他著「反射率関数の非線形近似化(Non-linear approximation of reflectance functions)」、Computer Graphics (SIGGRAPH '97 Proceedings), pp. 117-126, August 1997の概念化コサインローブ(generalized cosine lobes)の和によって各ポイントに記述されている。

#### 【0084】アルベドマップの構築

図示し、上述してきた実施形態では、拡散反射に起因する、空間的に変化する反射率を記述していることが要求されるアルベドマップは、制御された照明の下で撮影された顔のデジタル写真シーケンスを使用して測定されていた。

【0085】図11は、デジタル写真または画像を取り込むために使用された例示システムを示している。図示のシステムでは、デジタルカメラ1100が装備され、複数の光源を含んでいる。そのいくつかの例は1102、1104で示されている。垂直ポラライザ1106、1108、および1110の形体をした偏光フィルタが装備され、光源とカメラレンズを被覆しているため、鏡面反射が抑止され、拡散反射成分だけが画像に残るようにしている。この例では、被写体は、マシン認識可能ターゲットがプリントされた帽子1112を着用し、頭のポーズをトラッキングするようにしている。カメラ1110は、被写体が回転している間静止したままになっている。唯一の照明は、カメラ付近の測定個所に置かれた光源1102、1104から得ている。漏れた光からの間接的反射を減らすため、黒のバックドロップ(背景幕)が使用されている。

【0086】カメラと光源のロケーションは分かっているので、面法線、放射照度(irradiance)、視点方向、およびテクスチャ空間内の対応する座標は、各画像のピクセルごとに標準的光線トレーシング手法を使用して計算することができる。理想的なランベルト(Lambertian)反射が観察されるとの想定の下では、ランベルト反射率は、テクスチャ空間内の特定のポイントについてこの情報から計算することができる。この計算は、1つの写真に含まれるピクセルごとに繰り返され、その総和が得られると、画像がテクスチャ空間内に投影され、光源

に起因する計算された放射照度で除すると、面を横切る拡散反射率のマップが得られる。図12に示すように、2つの写真(A)、(B)がテクスチャ空間内に投影され、照明が補正されている。実際には、この投影は、外側ループがテクスチャマップ内のすべてのピクセルを通過するように繰り返される逆マッピング(reverse mapping)によって行われ、特定のテクスチャピクセルに投影される画像エリア全体の平均をとるために、確率的スーパーサンプリング(stochastic supersampling)が使用されている。

【0087】単一の写真からのアルベドマップは、面の一部だけをカバーしており、その結果は、グレージング(grazing)角度が小さいとき最良になっている。従って、顔全体のシングルアルベドマップを作成するために、すべての個別マップの加重平均が計算されている。重み付け関数は、そのビジュアル例が図13(A)、(B)に示されているが、面に対してほぼ法線をなしている方向から表示および/または照射されるピクセルには高い重みが与えられるように選択され、表示が照射のどちらかが極度にグレージングする前にゼロに降下するようになっている。図示の例では、次の関数が使用されている。 $(\cos \theta_i \cos \theta_e - c)^p$ 、ただし、 $c = 0.2$ ,  $p = 4$ 。

【0088】特定のテクスチャピクセルのアルベドを計算する前に、ピクセルが可視であり、適当に照射されていることが確かめられる。複数の光線は、そのピクセル上のポイントから光源上のポイントまでとカメラポイントまでトレースされ、そのピクセルには、可視性と照明がゼロであるか、部分的であるか、完全であるかのマークが付けられる。光源のサイズを見積もるとき、重大なエラーを避けるのが賢明である。完全に可視であり、少なくとも1つの光源によって完全に照射され、どの光源によっても部分的に照射されないピクセルのアルベドだけが計算される。このようにすると、部分的に隠されたピクセルおよび全影または半影領域にあるピクセルは使用されることがない。

【0089】これらの測定に意味をもたせるためには、若干の較正が必要である。カメラの転写曲線は、Debevec他著「写真からの高ダイナミックレンジ放射輝度マップの回復(Recovering high dynamic range radiance maps from photographs)」、Computer Graphics (SIGGRAPH '97 Proceedings), pp. 369-378, August 1997に記載されている方法を使用して較正されている。光/カメラシステムのフラットフィールド応答は、大きなホワイトカードの写真を使用して較正されている。レンズのフォーカルレングスとひずみは、Zhang著「カメラ較正のためのフレキシブルな新手法(A flexible new technique for camera calibration)」、Technical Report MSR-TR-98-71, Microsoft Research, 1998に記載されている手法を用いて較正されている。全体スケールファクタは、

既知反射率の参照サンプルを使用してセットされている。光源強度の画像間の変化が考慮の対象となっているときは、すべての画像に参照サンプルを組み込むことによって制御が行われている。

【0090】上記プロセスから得られたテクスチャマップは、顔全体にわたる詳細なカラー変化を自動的に取り込む上で良好な働きをしている。さらに、校正帽子を定位置に保持するために使用されるストラップは可視になっている。これらの問題は、画像編集ツールを使用し、ブランクエリアに近傍テクスチャまたは均一カラーを埋め込むことによって解消されている。

【0091】図14(A), (B)と図15(A), (B)は、編集前のアルベドマップと編集後のアルベドマップを、比較目的で示したものである。アルベドマップから妥当な結果が得られない個所がどこであるかは、面が十分に観察されない個所(例えば、あごの下)であるか、あるいは面が余りに複雑な形状をしているため画像と共に正しくスキャンし、記録できない個所(例えば、耳)であることは明らかである。これらの個所は、どちらも、リアル感のある外観を得るためにアルベドマップからのテクスチャを必要としていない。第1の個所は視覚的に目立たないからであり、第2の個所は、ジオメトリがビジュアル的に詳細になっているからである。そのため、これらの個所を編集しても、最終的レンダリングの外観はほとんど影響されない。

【0092】図16～図18は、付随するビデオからの静止フレームを使用した顔モデルのいくつかの、異なる様相を示す図である。1行目(図16の(A),

(B), (C))には、いくつかの角度からの顔が示されている。そこには、アルベドマップと測定されたBRDFが、任意の角度から見た顔全体にわたる皮膚の外観とそのカラー変化をどのように取り込むかが示されている。2行目(図17の(A), (B), (C))は、グレーディング角度での強い鏡面反射を含む、リムとサイドの照明の効果を示している。なお、光源は、この行の3画像のいずれも、同じ強度であり、顔から同じ距離に置かれている。反射率の方向変化は、親しみのある照明効果がレンダリングに現れることに貢献している。3行目(図18の(A), (B), (C))では、表情の変形が顔に加えられても、顔は通常の表情の動きでは自然のように見えることを示している。

【0093】図19は、ここで説明している実施形態に従ってアルベドマップを作成するための方法のステップを記載しているフローチャートである。この方法は、適切なハードウェア、ソフトウェア、ファームウェアまたはこれらの組み合わせで実現することが可能である。ここで説明している実施形態では、この方法は、図11を参照して説明してきたシステムと同じシステムで利用されるものとして、ソフトウェアで実現されている。

【0094】ステップ1700で、被写体を照射するた

めに使用できる1つまたは2つ以上の偏光光源が準備される。光源の例は上述した通りである。ここで説明している実施形態では、光源は、被写体の顔反射率の鏡面反射成分が抑止または除去されるように選択されている。ステップ1702で、被写体の顔は光源で照射される。ステップ1704で、被写体を回転させ、その間に一連のデジタル写真または画像が取り込まれる。ステップ1706で、テクスチャマップ内のピクセルごとに、面法線、放射照度、視点方向およびテクスチャ空間内の座標が計算される。この計算は公知のアルゴリズムを使用して行われる。ステップ1708で、画像のテクスチャ空間内の特定ピクセルに対するランペルト反射率が計算される。これにより、ピクセルのアルベドが得られる。ステップ1712で、追加のピクセルがアルベドマップに残っているかどうか判断される。残っていれば、ステップ1712で次のピクセルが得られ、ステップ1708に戻る。追加のピクセルがアルベドマップに残っていなければ、ステップ1714で、追加のデジタル画像が残っているかどうか確かめられる。追加のデジタル画像が残っていれば、ステップ1716で次のデジタル画像が得られ、ステップ1706に戻る。追加のデジタル画像が残っていなければ、ステップ1718で、各画像の個別アルベドマップの加重平均が計算され、顔全体のシングルアルベドマップが作成される。この加重平均処理がどのように行われるかの具体例は、Marschner著

「コンピュータグラフィックスの逆レンダリング(Inverse Rendering for Computer Graphics)」、PhD thesis, Cornell University, August 1998に記載されている。

【0095】図20は、シングルピクセルのアルベドを計算する方法のステップを記載しているフローチャートである。この方法は、適切なハードウェア、ソフトウェア、ファームウェアまたはこれらの組み合わせで実現することが可能である。ここで説明している実施形態では、この方法はソフトウェアで実現されている。ステップ1800で、あるピクセルが与えられているとき、そのピクセルが完全に可視であるかどうか判断される。ピクセルが完全に可視でなければ、ピクセルのアルベドは計算されない(ステップ1804)。ピクセルが完全に可視であれば、ステップ1802で、ピクセルが少なくとも1つの光源で完全に照射されているかどうか判断される。ピクセルが少なくとも1つの光源で完全に照射されていないければ、ピクセルのアルベドは計算されない(ステップ1804)。ピクセルが少なくとも1つの光源で完全に照射されていれば、ステップ1806で、ピクセルがいくつかの光源で部分的に照射されているかどうか判断される。そうであれば、そのピクセルについては、アルベドは計算されない。ピクセルがいくつかの光源で部分的に照射されていないければ、ステップ1808で、そのピクセルのアルベドと重みが計算される。この重みは、あとで個別マップをまとめて、その平均を

10

20

30

40

50

とるとき使用される。従って、上述したように、アルベドが計算されるのは、完全に可視であり、少なくとも1つの光源で完全に照射され、いずれかの光源で部分的に照射されていないピクセルについてだけである。このようにすると、部分的に隠されているピクセルと全影または半影領域にあるピクセルは使用されることがない。

【0096】以上、説明した顔の特徴をアニメーション化する方法および表情変換のための方法を実現するためのアプリケーションプログラムは当業者であれば、本明細書の説明に基づいて、当業者であれば作成することが可能であろう。

【0097】アプリケーションプログラムはCPUが実行可能な読取り実行可能な命令で作成すればよい。また、上記アプリケーションプログラムを記憶しておく記録媒体は、メモリ、ハードディスク、周知の携帯用記録媒体に記憶すればよく、これらの記録媒体は1つまたは2つ以上でもよい。たとえば、アプリケーションで使用するデータ、たとえば、コードブックやトレーニング表情を定義したデータ等はCDROMに記憶しておき、これらのデータを使用するアプリケーションプログラムはハードディスクに記憶しておくというように異なる記録媒体にアプリケーションプログラムとデータとを記憶させてもよい。また、同一の記録媒体に記録させてもよいこと勿論である。

【0098】さらにアプリケーションプログラムを実行するコンピュータは1台でもよいしアプリケーションプログラムで規定した処理を複数台のコンピュータで分散処理してもよい。

#### 【0099】

##### 【発明の効果】結論

上述してきた実施形態によれば、顔をリアル感のある顔にモデリングし、レンダリングするという高い基準を満たしてから、人間の顔と同じように親しみのある画像が信頼できる形で表示できるようにするという難題を解決するシステムと方法が提供されている。このアプローチの考え方は、顔モデルが実際に現実の顔に似るように、可能な限り測定を利用することにある。顔のジオメトリは、異なる顔に共通して統一した接続性と対応性をもつ、変位でマッピングされた細分面で表されている。反射率は、従来による人間の皮膚のBRDF測定から得られ、その測定が新規な測定と併用されて、複数の視点（ビュー）が拡散反射率を表す、単一の照明補正テクスチャマップになるように結合されている。動きは、上述した動き取り込み手法によって得られ、その動きが、改良された変形方法を使用して顔モデルに適用されることによって、シェーディング面に適した動きが得られるようにしている。レンダリングのリアル感、現実の顔のジオメトリ、動きおよび反射率を物理ベースのレンダラ（physically-based renderer）で使用することによって大幅に向上されている。

【0100】以上、構造上の特徴および/または方法ステップからとらえて本発明を説明してきたが、以上の説明から理解されるように、請求項に明確化されている本発明は、上述した具体的特徴または方法ステップに必ずしも限定されるものではない。むしろ、具体的特徴および方法ステップは、請求項に記載の発明を実現する、好ましい実施形態として開示されたものである。

#### 【図面の簡単な説明】

【図1】本発明の実施形態を実現する上で使用するのに適している汎用コンピュータを示すブロック図である。

【図2】被写体の顔の構造情報と反射率情報とを同時に取り込むために利用できるシステムを示す構成図である。

【図3】本発明の実施形態に従って構造情報と反射率情報を取り込むための例示方法を記載しているフローチャートである。

【図4】本発明の実施形態による例示コードブックと変換関数を示す説明図である。

【図5】本発明の実施形態による表情変換プロセスを示すフローチャートである。

【図6】本発明の実施形態のある種の原理が採用可能である例示システムを示すブロック図である。

【図7】（A）～（C）は本発明の実施形態による例示表情変換を示す例示カラープレートのコレクションを示す図である。

【図8】同一細分制御メッシュを、顔が異なるごとに変位された細分面にマッピングするプロセスを示す図である。

【図9】（A）および（B）は面適合過程で特徴の対応づけを行うために利用される例示制約を示す図である。

【図10】本発明の実施形態による面適合方法のステップを記載しているフローチャートである。

【図11】本発明の実施形態に従って顔のアルベドマップを作るために採用できる例示システムを示す構成図である。

【図12】（A）および（B）はテクスチャ空間内に投影され、照明が補正された2つの写真の例示アルベドマップを示す図である。

【図13】（A）および（B）は図12に示す写真に対応する例示重み付け関数を示す図である。

【図14】（A）および（B）は2つの異なるデータセットに対する2つの完全アルベドマップを示す図である。

【図15】（A）および（B）は編集したあとの図14のアルベドマップを示す図である。

【図16】（A）～（C）は異なる照明条件の下で異なる向きにレンダリングされた顔モデルのコレクションを示す図である。

【図17】（A）～（C）は異なる照明条件の下で異なる向きにレンダリングされた顔モデルのコレクションを

示す図である。

【図 18】(A)～(C)は異なる照明条件の下で異なる向きにレンダリングされた顔モデルのコレクションを示す図である。

【図 19】本発明の実施形態に従ってアルベドマップを作成するための方法のステップを記載しているフローチャートである。

【図 20】本発明の実施形態に従ってシングルピクセルのアルベドを計算するための方法のステップを記載しているフローチャートである。

【符号の説明】

200 照明システム

202 カメラ

202a カメラのレンズ

204、206 光源

400 コードブック

402 トレーニング表情集合

406 変換プロセッサ

408 変換関数

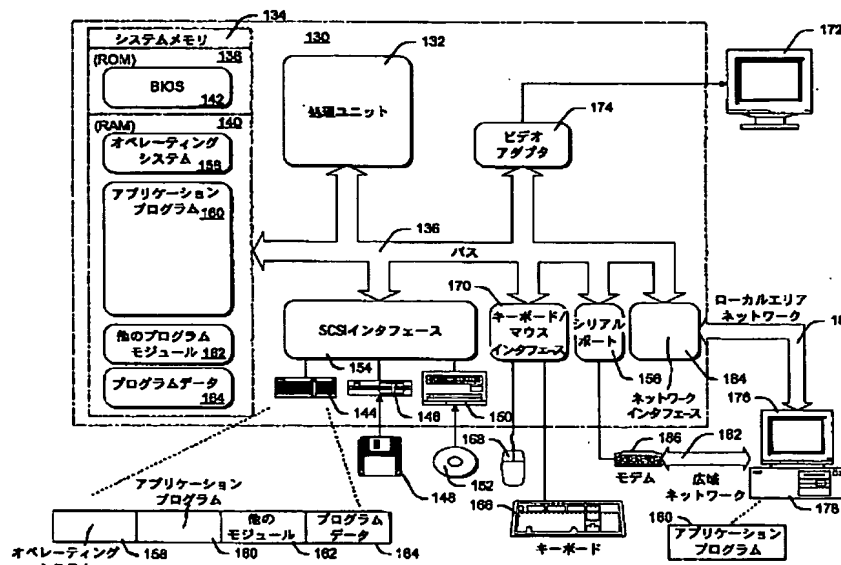
410 合成表情集合

602 トランスミッタ (送信側)

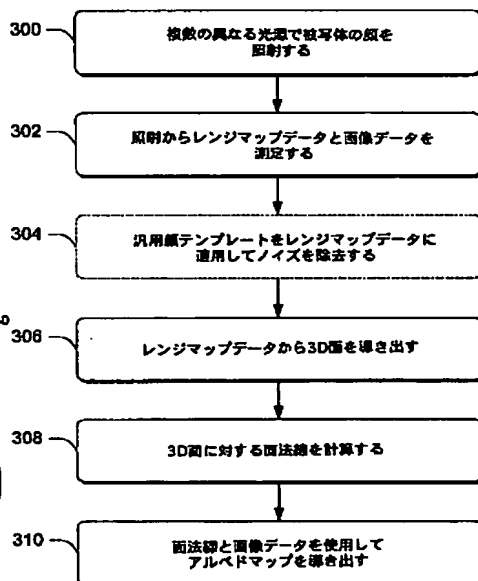
10 604 レシーバ (受信側)

606 再構築モジュール

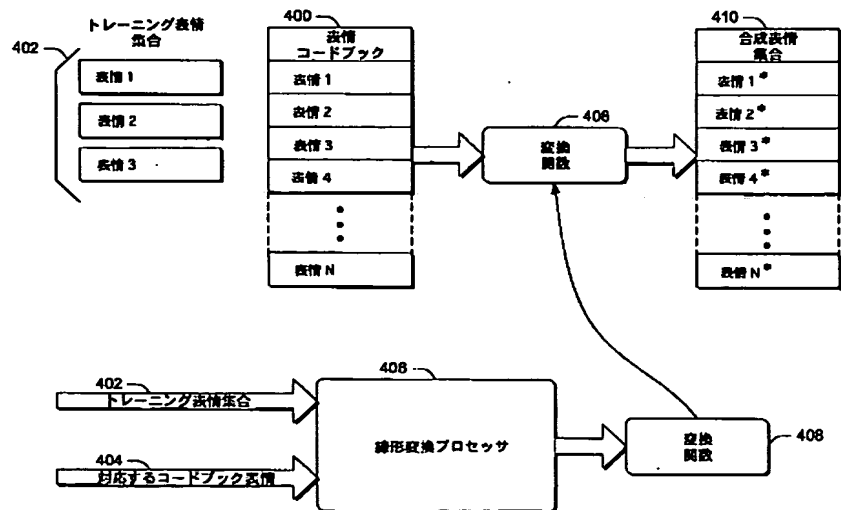
【図 1】



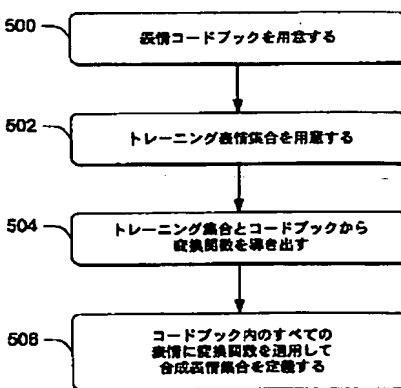
【図 3】



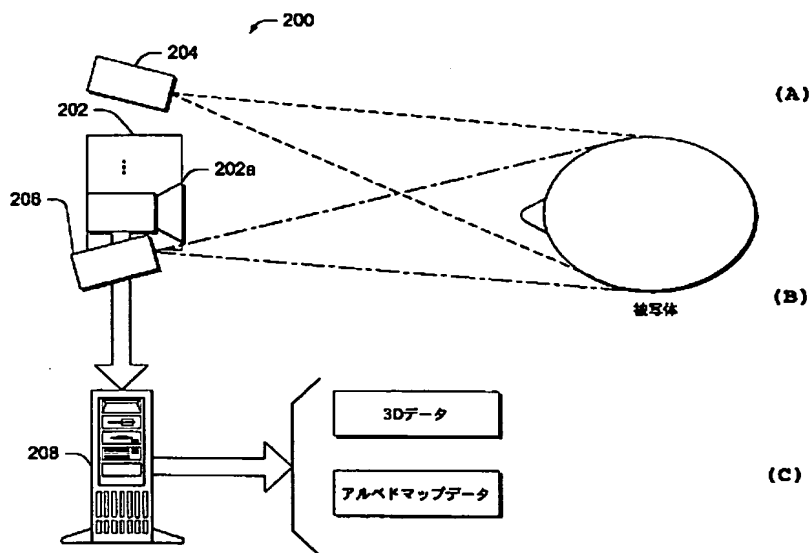
【図 4】



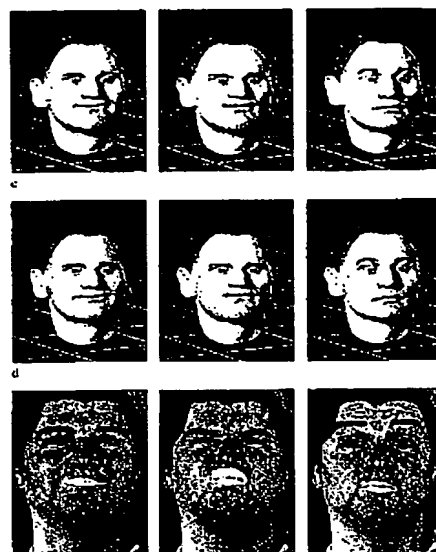
【図 5】



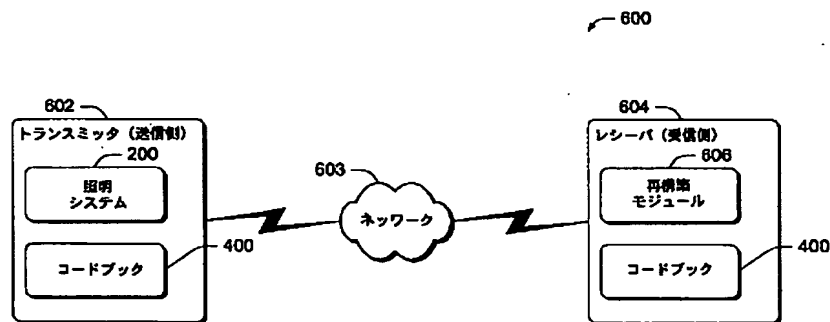
【図 2】



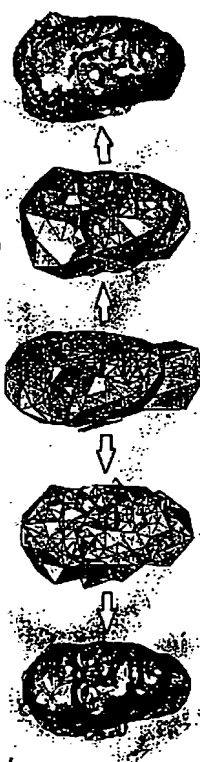
【図 7】



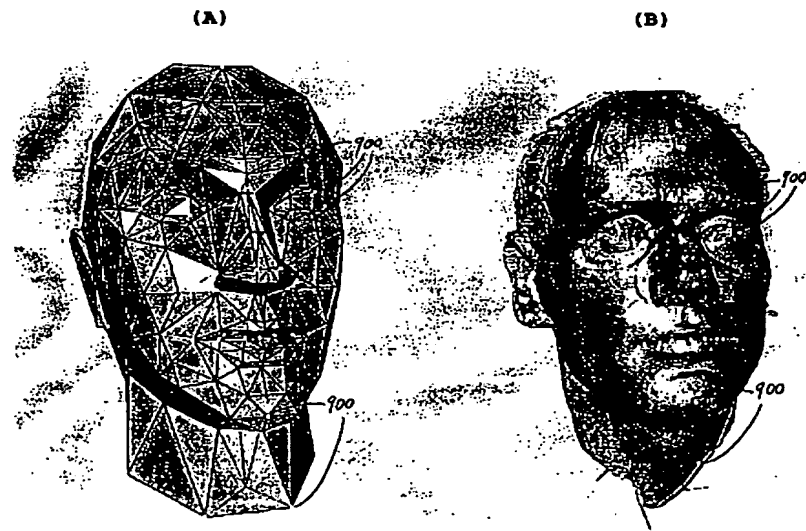
【図 6】



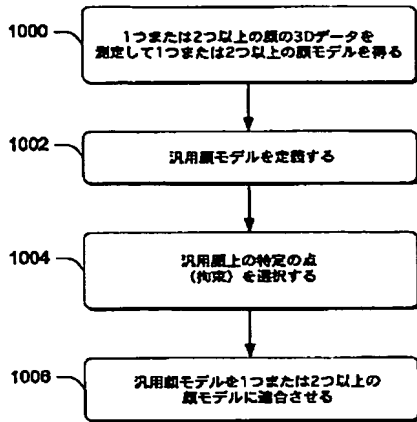
【図 8】



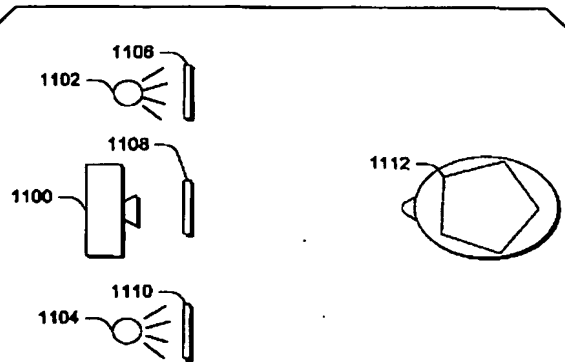
【図 9】



【図10】



【図11】



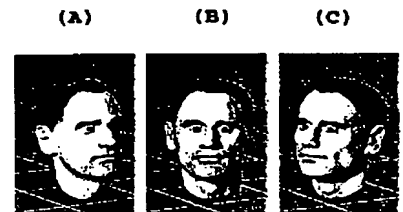
【図12】



【図13】

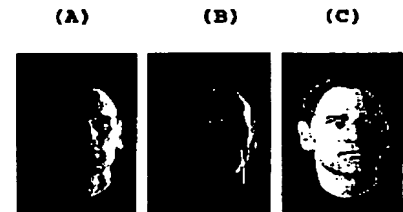


【図16】



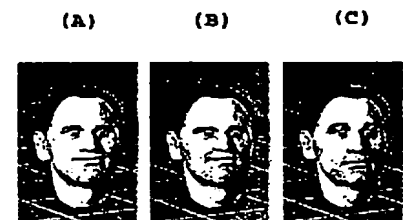
1行目

【図17】



2行目

【図18】



3行目



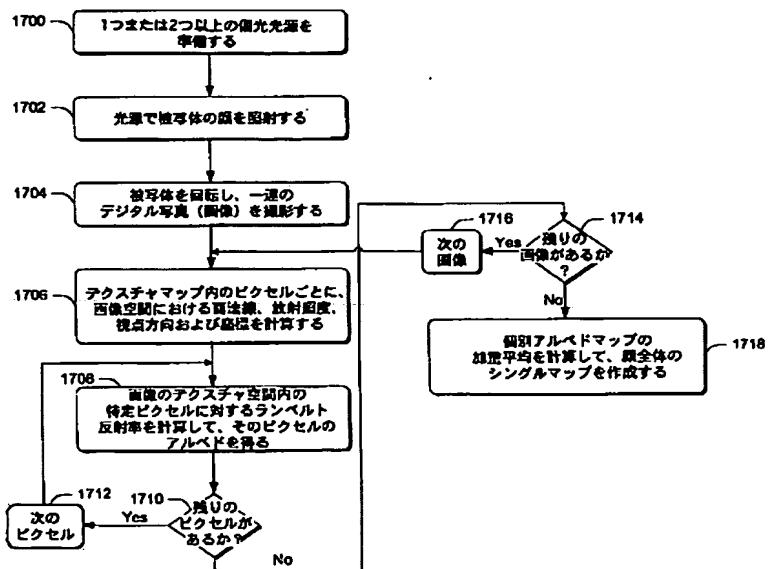
【図14】



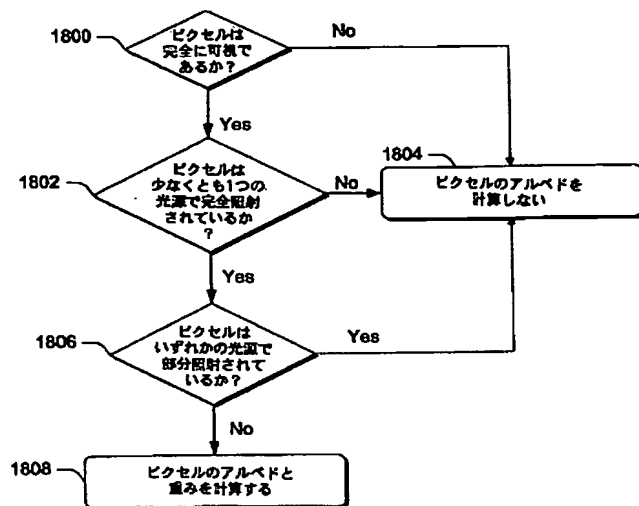
【図15】



【図19】



【図 20】



フロントページの続き

(72)発明者 ブライアン ケイ ジェンター  
アメリカ合衆国 98052 ワシントン州  
レッドモンド ノースイースト 41 スト  
リート 16725

(72)発明者 サシ ラグフバティー  
アメリカ合衆国 98052 ワシントン州  
レッドモンド 149 コート ノースイ  
ースト 6110

(72)発明者 カーク オリニク  
アメリカ合衆国 98052 ワシントン州  
レッドモンド 157 コート ノースイ  
ースト 13517

(72)発明者 シン ビン カン  
アメリカ合衆国 98052 ワシントン州  
レッドモンド ノースイースト 100 コ  
ート 18217

Fターム(参考) 5B050 AA08 AA09 BA08 BA09 BA12  
CA08 EA19 EA27 EA28 EA30  
FA02  
5B057 AA20 BA02 BA08 CA13 CA16  
CB13 CB16 CC01 CD11 CE08  
5L096 AA06 AA09 BA08 BA18 CA05  
CA17 DA02 FA66

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☒ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**